THE UNIVERSITY OF ALBERTA

PREDICTING TRAINING OUTCOMES FOR STUDENTS IN A TECHNOLOGICAL INSTITUTE

by

© Paulino I. Villagonzalo

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES

IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE

OF DOCTOR OF PHILOSOPHY

DEPARTMENT OF EDUCATIONAL PSYCHOLOGY

EDMONTON, ALBERTA

FALL, 1969

UNIVERSITY OF ALBERTA

FACULTY OF GRADUATE STUDIES

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies for acceptance, a thesis entitled "Predicting Training Outcomes for Students in a Technological Institute," submitted by Paulino I. Villagonzalo in partial fulfilment of the requirements for the degree of Doctor of Philosophy.

ABSTRACT

This study investigated whether a combination of nineteen predictor variables could differentiate between nine training groups of graduates from the Northern Alberta Institute of Technology. The purpose of the study was to develop a procedure for predicting training outcomes of students in a technological institute.

Discriminant analysis was used in determining the reduced space in which the separation of training groups was maximum. The person's centour score was used as a basis of determining his group membership. Various combinations of predictors and training groups were analyzed and the results of these classifications compared. The effectiveness of classification was verified from a stratified random sample of graduates which was set aside for validation. It was also verified from the criterion sample used in the analysis.

Classification of group membership was determined by the conditional and Bayesian rules of probability. Comparison of classification results were made in terms of number and corresponding percentages of correct classifications, close classifications, fairly close classifications, overall classifications, and misclassification.

Out of the five discriminant analyses carried out, analysis 2, involving eight training groups and nineteen predictors, showed the best result of classification. The discriminant weights obtained in this analysis were applied on a sample of non-graduates to verify the classification procedure employing a different group of students. The

result of the classification showed a reversed trend. This was another indication that the prediction scheme developed in this study could be a useful tool for counseling.

Reducing the number of training groups by combining two of them showed an improvement in correct classification, but not always in accuracy. When the analysis was limited to highly weighted predictors, correct classifications were found to be lower than when 19 predictors were used. Twelve predictors, which were readily available to counselors at NAIT, could probably have served the purpose of counseling although with a lesser degree of effectiveness.

Psychological meanings were derived from the distribution of scores of the graduate sample in a 3-dimensional discriminant space. The first function was interpreted as measuring a mechanical attribute, the second as an intellectual attribute, and the third as a verbal-enterprising attribute.

The analysis carried out to determine whether 19 predictors could differentiate between two outcome categories, showed that it was not possible to differentiate between graduates and non-graduates. One reason advanced was that this study used a restricted group.

The study demonstrated that 19 predictor variables could be analyzed to develop a workable prediction scheme for purposes of counseling in a technological institute.

ACKNOWLEDGEMENT

The writer gratefully acknowledges the advice and assistance received from Dr. V. R. Nyberg, Dr. J. E. Bicknell, and Dr. T. O. Maguire, University of Alberta, who offered many constructive comments and suggestions on this thesis. Their interest and concern about the program of activities of the writer at the university are heartily appreciated.

Acknowledgement is also extended to the Division of Educational Research Services of the university for the assistance given with regards to statistical analysis, particularly to K. Bay, a staff member of the Division, who prepared and checked the FORTRAN programs used in the study and to E. Romaniuk, also staff member of this Division, who helped the writer in preliminary editing.

Sincere thanks are extended to the administrative staff of the Northern Alberta Institute of Technology, to S. Checkley, Head of the Guidance and Counseling Services of the institute, and to the Examination Branch, Alberta Department of Education for facilitating the gathering of the data.

The writer is also grateful to the Canadian International Development Agency, the Department of Foreign Affairs and the Department of Education of the Philippines, to R. Y. Mendoza, Director of Vocational Education, Manila, and to M. S. Bonilla, Superintendent, Cebu School of Arts and Trades, Cebu City, Philippines for making possible the realization of the degree program of the writer here in Canada.

TABLE OF CONTENTS

CHAPTER PAGE

LIST OF TABLES

TABLE PAGE

TABLE PAGE

TABLE PAGE

LIST OF FIGURES AND CHART

CHAPTER I

INTRODUCTION

Super (1954) viewed vocational counseling in the United States and Canada as "human development" while viewing it as "manpower utilization" in countries with disturbed economies. In the latter case its purpose is primarily a means of obtaining the needed supply and distribution of trained manpower. Human development on the other hand implies the development of the individual to the full maximum of his potentialities and to his own satisfaction. In the North American cultural framework vocational counseling is more extensive in scope than elsewhere. It places more demands on the counselor to apply various methods and procedures in assisting the student to attain that full measure of human development and satisfaction.

When a student comes to him for advice with regard to choosing an occupation the counselor studies the characteristics of the individual and looks over occupations in which others with characteristics like his were successful and satisfied. To secure this information which consists of antecedent-consequent relationships the counselor uses both the statistical and clinical approaches. As used here, statistical approach refers to test interpretation and prediction. The clinical approach refers to inferential, subjective, and nonquantitative prediction. The clinical approach will not be treated in this dissertation.

One of the most common statistical bridges the guidance counselor uses in his work is the test norm. This procedure involves a direct comparison of the individual's raw test score or converted score with

some table of norms. It is based on the assumption that the norm group represents a meaningful base for comparison and prediction. Goldman (1961) observed that the greatest danger in the use of norms is that more will be assumed for a normative statement than is warranted. Another danger is the unwarranted assumption that the higher the score, in comparison with the norm group, the better. In spite of their limitations, test norms have some practical uses in vocational counseling.

For many years the mainstay of test interpretation and prediction in vocational counseling has been the regression approach. In this procedure prediction equations are computed, based on the linear relationship between test scores and some criteria of success or satisfaction. The multiple regression technique is used to select the most effective combination of tests, determine their respective weights and assess the effectiveness of the composite predictor. By substituting a person's test scores in the multiple regression formula, his predicted criterion score can be determined and made use of in vocational counseling.

However, there are some difficulties with the use of the multiple regression approach in test interpretations and prediction. Rulon (1951) pointed out that it is almost always assumed in this approach that the person in question is a member of the group. This is so because the regression equation is developed on the basis of persons who have already entered a particular group. It cannot be denied that there are many selection factors involved before the individual becomes a member of the group. Cooley and Lohnes (1962) cited an example of chemistry majors who might all tend to be high on mathematics aptitude. By the multiple regression procedure the importance of mathematics aptitude

would not be reflected in predicting performance as a chemistry major. This might result in misclassification. With regression information the individual is more likely to appear to have the highest aptitude for the less difficult task hence the tendency for such an individual to enter or be assigned to less difficult tasks. Although the results of the regression calculation indicate relationships between predictors and criteria for individuals within groups it is lacking with regards to relationships between groups. In his work with classification and selection for Air Force personnel, Tiedeman (1951) pointed out a further difficulty in using regression analysis. The problem is in making criterion scales comparable from one job to the next.

In spite of these limitations, correlation and regression analysis do give useful information and have contributed to a large extent in the development of test interpretation and prediction and in the advancement of psychological knowledge. The approach is useful in predicting the group in which an individual will be most successful. It is also useful when individuals in a group are ranked according to their predicted ability to perform a particular job in order that a certain number may be selected from the top.

However, in personnel classification and vocational counseling one is faced with the question "To which group does an individual bear the greatest resemblance?", before asking the question, "How well will the individual perform if assigned to the group he most resembles?". Surely, the answer to the first question would provide vocational counseling with an effective means of uncovering promising careers for the individual and assisting him to make an appropriate choice. This

may enable the individual to set his goal toward the most rewarding vocation for him as well as to attain the purposes of vocational counseling which Super termed as "human development". Yet while the first question is basic in helping an individual make proper choices, the use of an appropriate statistical technique such as discriminant analysis is rarely if ever used in school guidance. The reason for this situation is, perhaps, the need for well-informed personnel to carry out possible research applications. Another reason is that the determination of the function requires voluminous calculations when the analysis involves three or more variables and groups. However, with the aid of the medium speed electronic computer, which is becoming available in many school systems, the above reasons may now be overcome.

Multiple discriminant analysis originated with Fisher. It is widely used in taxonomic scientific investigations such as dating a series of Egyptian skulls, distinguishing between two forms of black locust trees, and classifying individuals into three Indian castes. These earlier investigations mentioned above were carried out by M. Barnard, Fisher, and Rao respectively, (Tiedeman, 1951). Since then this technique has been used in many scientific classification and selection problems. Its application to psychology is relatively recent and not widely developed. Current basic statistical textbooks in psychology and education rarely provide a complete explanation regarding methods of calculating the discriminant function as opposed to the widely known multiple regression analysis.

The early investigators who conducted studies using discriminant analysis most often concerned themselves with demonstrating its effectiveness and discriminating power in psychological classification problems.

Sufficient evidence has now accumulated indicating the usefulness of the multiple discriminant function as a statistical tool for certain psychological investigations. The recent direction of studies in this regard has shifted to investigation of predictor variables and group categories that may be useful in identifying special talents, in determining vocational choices, and in classification and selection of personnel. However, by and large the movement along the continuum from statistical and psychological knowledge to educational practice has been a slow one. Further studies are still needed to identify effective predictor variables appropriate for particular situations such as those in vocational guidance and counseling, student selection, and manpower training and utilization.

CHAPTER II

THE PROBLEM

Each year an increasingly large number of students of ages 16 years or over, who had completed the eleventh and twelfth grades in high schools in the province of Alberta and other Canadian provinces, applied for enrollment at the Northern Alberta Institute of Technology to pursue various technical, business, and vocational courses. These applicants were admitted primarily on the basis of grade level completed, marks earned, and priority of application.

According to the 1966-67 school calendar[1] the institute offered an articulated technology program in three patterns of studies. The three-year pattern of studies consisted of Year A, B, and C. Entrance requirements for admission into Year A of this pattern were: 67 Alberta high school credits (more or less a completion of Grade XI) with at least a "B" standing in Mathematics 20 or 22, Science 20 or 22, and English 20. However, preference was given to applicants having additional Grade XII credits particularly in Mathematics 30 or 32. Year A was a pre-technology program consisting of Grade XII mathematics, physics, English, plus the applicable vocational subject. This program was, in some measure, a making up of the deficiency of the student for not completing Grade XII and earning an Alberta high school diploma.

---

[1] There have been some changes in the school calendar of the institute since then. However, these changes, up to the date of this writing, did not significantly affect the discussion above.

One of the two-year patterns of study consisted of Years B and C. The entrance requirements for admission to Year B of this pattern were: an Alberta high school diploma (i.e. a completion of Grade XII) with at least a "B" standing in Mathematics 30 or 33, thirty-five or more credits in one of the articulated vocational high school subjects with at least a "B" standing in the final year and credit in Physics 30 or 32, or a successful completion of Year A at the institute in Edmonton or Calgary[2].

Another two-year pattern consisted of Years AB and C. This pattern accommodated students coming from the high school academic program. The entrance requirements for admission to Year AB in this pattern were: an Alberta high school diploma with at least a "B" standing in Mathematics 30 or 32, Physics 30 or 32, a credit in English 30 or 33 and a minimum overall high school average of 55%. Year AB was given on an extended term of 10 months with 33 hours per week of instruction. It was felt that this extension of study time was sufficient to cover Years A and B considering that those who were admitted into this pattern had high academic standing.

The articulated technology courses offered under these patterns were: air conditioning and refrigeration, electrical, architectural, drafting, electronics, exploration, instrumentation, industrial production, and telecommunication technologies. Figure 1 below showed the various routes under this program.

---

[2] The province of Alberta operates two institutes of technology, the Northern Alberta Institute of Technology in Edmonton (NAIT) and the Southern Alberta Institute of Technology in Calgary (SAIT).

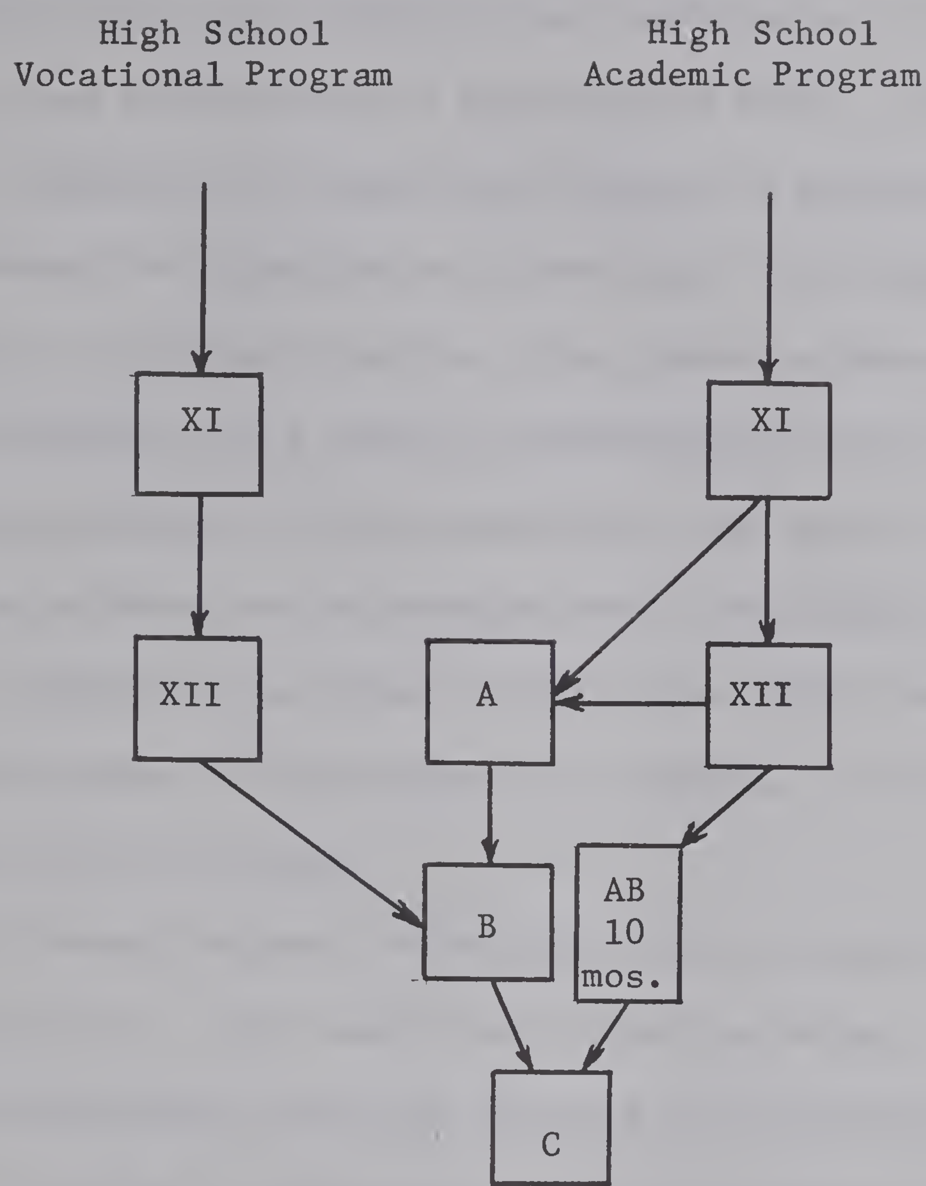

Fig. 1

Diagram showing the three patterns of studies of the articulated technology program

The institute also offered a non-articulated technology program. These were technologies with specific entrance requirements and length of training. The non-articulated technologies which normally required two years to complete were: chemical, civil, surveying, gas, materials,

plastics, dental laboratory, forest, heavy duty equipment, photographic, medical laboratory, and medical X-ray technologies. The entrance requirements in these non-articulated technologies were: a high school diploma and a "B" standing in a high school subject or subjects considered most requisite for a particular technology. For example, in chemical technology the applicant must be a high school graduate with at least a "B" in chemistry and a credit in Mathematics 30 or 32. The medical laboratory and medical X-ray technologies had special requirements. Applicants to these two technologies must have Alberta (senior) matriculation. Preference was given to the student with credits in English 30, Social Studies 30, Mathematics 30, Language 30, and two of Chemistry 30, Physics 30 or Biology.

The non-articulated technologies which normally required one year to complete were: radio and TV and dental assisting. Applicants to these technologies must have obtained 67 high school credits and a "B" standing in Mathematics 20 or 22, Science 20 or 22, and English 20.

The business and vocational courses which normally required two years to complete were business administration, commercial cooking, dietary service technician, distributive technology, electronic data processing, and secretarial technology. Courses normally requiring one year for completion were banking program and office machine mechanics. The minimum requirement in the banking program, business administration, distributive technology, and secretarial technology was a high school diploma. Preference was given to applicants who had 50% in Grade XII English and Grade XI mathematics. Commercial cooking required 35 Alberta high school credits. Applicants to the dietary service technician

course must have completed Grade XI with at least 50% grade in Grade XI mathematics, science, and English. In the electronic data processing course the applicant must have had a high school diploma with a "B" in mathematics 30 or 32 and Grade XII English. Applicants for the office machine mechanics course were required to complete Grade XI only.

To be considered for admission to the institute the applicant must file an application form stating such information as the course desired, age, height, weight, disabilities, church affiliation, marital status, courses previously taken at the institute if any, high school subjects completed and corresponding marks in Grades X, XI, and XII, and related work experiences. As a check an official transcript of all high school marks was attached to the application as soon as this became available.

The selection of students who were finally accepted for enrollment was decided on the basis of the prescribed requirements in the school calendar. If there were any doubts as to the capabilities of the student or his success in a certain course the student was referred for interview with the head of the department concerned and/or a guidance counselor and their recommendations were combined with the other requirements in considering the final admission of the student.

As may be gleaned from this procedure of student selection reviewed above, the course choice of the applicant could be attributed to opportuneness, that is, both his personal wish and the institution's will were met. It could be attributed in some degree to the results of the interview with the department head and/or a guidance counselor and perhaps it could be attributed in some degree to the vocational and academic experiences of the student while in high school, to his work

experiences outside the school, and to his maturing awareness of his abilities and aptitudes.

Considering the host of technologies and courses which a student could choose from, one of the most important responsibilities of the department heads and guidance counselors who assisted the applicant in making an appropriate choice was to judge the courses for which the student had an aptitude. To discharge this responsibility required some basis of comparing the characteristics of the student with groups to which he might probably best belong. The entrance requirements prescribed by the institute indicated that some assumptions were made concerning a number of abilities of the individual and the different degrees and combination which were required for success and satisfaction in the various technologies and courses. If a number of characteristics of the student were observed at the early part of admission or before admission to his course and then set aside until after his graduation, one might be able to study how this information could be combined in making a recommendation of enrollment to a certain course for incoming students to the institute. Such study would give substantial assistance to the applicant who was making the choice as well as minimize misclassification of placement on the part of the institute.

During the admission period the basic problem in helping the student applicant make an appropriate choice was to provide a probable answer to the question, "Which group did he resemble most?", before one answered the question "How well will the individual performed if assigned to the group he most resembled?". In other words it was knowing whether the applicant most resembled the graduates of either electronics

or drafting or business administration or some other course. To provide better assistance to students studying at the Institute, with regard to choice of courses, it was felt that there was a need to develop some means of answering the first question above.

No doubt the many problems like the one cited above had been in the minds of the guidance and counseling personnel of the institute for some time. During the early months of the 1966-67 school term they administered the Differential Aptitude Tests and the Lorge-Thorndike Verbal and Non-verbal Intelligence Tests to all students enrolled for the first time in the various technologies, business, and vocational courses for purposes of securing information about students that might be helpful in vocational counseling. Raw scores of these tests were converted into percentile scores and intelligence quotients respectively using the tables of norms provided in the test manuals. To a certain extent these results gave some information useful in vocational counseling at the institute. However, as stated earlier, norms do not go very far for purposes of prediction especially if the sample was not comparable with the norming groups.

Studies conducted by Bennett, Seashore, and Wesman (1959) on the predictive qualities of the Differential Aptitude Test revealed important profile differences among high school students who entered diverse occupational and educational careers. A longitudinal study by Seashore (1959) showed that the level of the career endeavor of students was related to their level of aptitudes and that the job level within an occupational area was related to the scores on the most relevant sub-tests. According to the authors the Differential Aptitude Tests were not developed to isolate simple, pure abilities but rather to measure complex abilities

in relation to job families and curricula. However, a study by Guilford, Hoepner, & Petersen (1965) indicated that the combination of four DAT subtests did about as well as factored battery tests among several standard tests compared in their study. The Lorge-Thorndike Intelligence Tests were measures of abstract intelligence expressed in verbal symbols on one part and in pictorial, diagramatic, and numerical symbols on the other. Studies by Lorge & Thorndike (1954) had shown high correlations of these tests with educational achievement tests and school marks. Not many studies had been done with these tests on prediction over a period of time. Group mental tests, according to Cronbach (1960), predicted success on a great many jobs although in some jobs, special abilities were much more important than general ability. A review of literature on prediction of success of workers made by Ghiselli in 1955 substantiated this observation and indicated the importance of group mental tests in vocational guidance. With certain types of students and a school setting such as the one at Northern Alberta Institute of Technology these tests mentioned above had not been studied to determine the prediction of membership of individuals among different groups.

By the end of the 1967-1968 school term, two years later, the students who took those tests had graduated from the various technologies, business, and vocational courses. Those who started in Year A had either successfully completed Year B of their respective courses or had failed some of their subjects. Others had either withdrawn from the institute before they were able to complete their courses, transferred to another course and were continuing their studies in the institute, or did not graduate due to failure in some subjects.

Each year there were many applicants coming from the various high school programs who elected to pursue certain courses, who were admitted to the institute on the basis of grade level and marks earned on certain subjects, and who after the end of one or two years appeared in various training categories. However, very little information was provided to the entering student about how his aptitudes and high school achievements compared with those students who graduated or made satisfactory progress in the various courses.

The problem of this study was, therefore, to develop a procedure for interpreting the scores on the Lorge-Thorndike Intelligence Test, the Differential Aptitude Tests, the Grade Nine Alberta Departmental Examinations, School and College Ability Tests Level 3, and the high school averages for predicting the training outcomes of graduates of the Northern Alberta Institute of Technology.

So far, a review of the literature in this area showed that no study had been made on the variables stated above in predicting training outcomes of students in vocational and technical institutes. Terminal schools like the Northern Alberta Institute of Technology were a special type of institution which provided technical, business, and vocational training to a certain type of student who later make up the considerable bulk of the main work force of the nation and who no longer enter the higher institutions of learning. Very little research information was known about this type of student.

Theoretical framework

Tyler (1959) in her theoretical analysis toward a workable psychology of individuality pointed out that the emphasis on the

traditional measurement approach based on dimensions or trait continua might not likely resolve the prediction of human behavior. Some evidence had accumulated which indicated the need for the use of a kind of measurement which would characterize the person's customary pattern of choice and organization. One impressive piece of evidence which supported this contention was the 20-year follow-up study of the Strong Vocational Interest Blank (1955) based on letter grades as predictors of a special kind of criterion - that of remaining in the original occupation versus shifting to another. These letter grades carried different meaning from customary trait measurement. An "A" would signify that the person's customary pattern of choice and organization was like the choice pattern characteristics of persons in a certain occupation.

Cooley (1964) in a similar vein explained that the basic proposition was that different plans were appropriate for different people. The problem, was to develop a method for dealing with the relationships between types of plans and kinds of people.

From factor theory it was possible to deal with the taxonomic problem of plans and people. In conceptualizing human behavior in factor theory, personality had its locus in an $\underline{m}$-dimesnional space. The location of the individual's personality in this space was determined by the total pattern of the $\underline{m}$-behavioral measures which were available for that individual. Personality, as used here, referred to all behavior and included intellectual functioning as conceptualized by Guilford (1959). Individuals with similar characteristics tended to occupy similar regions in this $\underline{m}$-dimensional space. This meant that individuals with similar personalities would have a tendency to make a similar occupational choice. In this study occupational choice was

based on personality and not on things like occupation of the father, economic status, religion, etc. If we were able to determine the regions of the groups of individuals who had made particular choices, providing that these regions were distinct with respect to these personal characteristics and that the individuals were drawn from the population for whom each region was constructed, it would be possible to estimate the probability of the choice of another person with a similar pattern of personality.

Some studies suggested that different levels of aptitude were required of different educational and occupational levels. Some studies suggested that individuals developed a certain awareness of these differing degrees of abilities and aptitudes at some age level approximately during their senior or high school graduation. Also, one's choice of education and occupation was influenced to a certain extent by this maturing awareness of abilities and aptitudes. Taking cognizance of these findings, this study postulated that the scores made by individuals on the Lorge-Thorndike Intelligence Tests, the Differential Aptitude Tests, the Alberta Grade Nine Departmental Examinations, the School and College Ability Tests Level 3, and the high school averages could be used in determining different regions in an $\underline{m}$-dimensional test space which certain groups tended to occupy. It further postulated that certain regions defined in the $\underline{m}$-dimensional space were distinct with respect to these characteristics and that a person's likelihood of membership in a certain occupational group could be predicted from his pattern of scores on these tests and high school averages.

Null hypotheses

This study tested or investigated the following null hypotheses:

1. There are no significant differences among group means in the sub-scores of the Lorge-Thorndike Intelligence Tests, the Differential Aptitude Tests, the Alberta Grade Nine Departmental Examinations, SCAT Level 3, and the high school averages of the groups of students classified as graduates and non-graduates on an *a priori* basis according to training outcomes.

2. The characteristics of the various training groups as measured by the nineteen predictor variables consisting of the sub-scores of the Lorge-Thorndike Intelligence Tests, the Differential Aptitude Tests, the Alberta Grade Nine Departmental Examinations, SCAT Level 3, and the high school averages are not different.

3. The dispersions of groups along the discriminant vectors as defined by the sub-scores of the Lorge-Thorndike Intelligence Tests, the Differential Aptitude Tests, the Alberta Grade Nine Departmental Examinations, SCAT Level 3, and the high school averages are not significant.

4. The sub-scores of these tests stated above and the high school averages do not predict probable membership of students to certain training groups as classified on an *a priori* basis according to training outcomes.

5. There are no psychological meanings or reasons that might be derived for group positions in the discriminant space based on the sub-scores of these tests stated above and the high school averages.

CHAPTER III

REVIEW OF LITERATURE

The main problem of this study was to determine the membership of individuals in training outcome groups on the basis of a combination of predictor variables. Discriminant analysis had been found most appropriate for such a problem. Travers (1939) demonstrated its use in distinguishing successful engineers from successful air pilots. He secured twenty samples from each of the two groups and used six test scores of intelligence, form relations, dynamometer, dotting, sensory-motor coordination, and perserveration. However, the purpose of his study was merely to show the possibilities of the discriminant analysis in psychological investigations. Garret (1943) used the same data as Travers' but utilized the scores of three tests only. His purpose was also similar to Travers'. His interest was in explaining and showing the calculation of the discriminant function in such a manner that psychologists who were not so mathematically oriented may be able to make use of it.

One of the early psychological investigations on determining group membership using the discriminant function was done by Baggaley (1947) of Harvard University. He wanted to study the relation between the scores obtained by 180 Harvard freshmen on the Kuder Preference Record and their fields of concentration. He classified the 22 fields of concentration into two groups: Group A consisted of those who chose the natural science area and Group B consisted of those who chose the social sciences and humanities. His findings showed that the Kuder scores provided a basis for differentiating among Harvard students who

proposed to concentrate in different academic fields. However, the uses of the findings for guidance was rather limited to the classification of two broad categories of academic training outcomes since Groups A and B lumped together several fields of concentration.

It was only in 1950 that a method was developed to use discriminant analysis with more than two groups. Working independently, Bryan (1950) at Harvard developed a method for the exact determination of the characteristic equation and latent vectors of a matrix with application to the discriminant function for more than two groups. At about the same time Rao (1952) in India also independently generalized the method to investigate differences among the multivariate distribution on any number of groups in problems of biological classifications. Using the scores of the Kuder Preference Record of freshmen at Harvard, Bryan demonstrated that his method could be used on a large scale of predictor variables and reduced the numerical process of calculating the discriminant roots and latent vectors.

Tiedeman and Steinberg (1951) investigated the difference between the regression analysis and discriminant analysis in a classification problem and showed the possibility of misclassification in the former. The data used were the sub-scores of the Differential Aptitude Tests. The tests were administered to 207 beginning ninth grade students in Waltham, Massachusetts who were allowed to choose from the various secondary curricula. At the end of their tenth grade most of the students had chosen either the college preparatory or the business curricula. By regression analysis using the tenth grade average as criterion, the results showed that if Waltham students had chosen the curriculum in which they were more likely to receive their highest mark, the college

preparatory curriculum would be the choice of all the students. However, when the grade average criterion scores were transformed to standard scores the two regression equations when recomputed showed that the business curriculum regression was higher than the college preparatory curriculum regression. This result showed that choosing the curriculum in which Waltham students were more likely to receive the highest mark did not imply appropriate choice of curriculum for each student. However, in a discriminant analysis, the Differential Aptitude Tests clearly distinguished between students who chose either the college preparatory curriculum or the business curriculum. Although the study was limited to only two groups it showed the effectiveness of the discriminant analysis for classifying groups as compared with regression analysis.

Tiedeman, Bryan, and Rulon (1953) investigated the utility of the Airman Classification Battery for assignment of airmen to eight air force specialities using the method of calculating the multiple discriminant function developed by Bryan. Since airmen were already members of the Air Force the problem was to assign each airman to a job which maximized satisfaction with the total assignment. The seventeen stanine sub-scores of the Airman Classification Battery AC-1 used in the study were: clerical key, mechanical key, crafts key, equipment operator key, radio operator key, word knowledge, arithmetic reasoning, dial and table reading, numerical operations II, aviation information, background for current affairs, electrical information, mechanical principles, general mechanics, tool functions, speed of identification, memory of landmarks. Assignment to a technical training school was based on the choices of the airmen as expressed during the career counseling interview within restrictions of aptitude indices, quota, and personnel supply. The eight

technical specialities selected for the study were: radio operator, clerk-typist, control tower operator, aircraft and engine mechanic, radio mechanic, weather observer, airplane sheetmetal worker, and radar mechanic. After two years of technical training, 6105 airmen who were considered successful for having received an average grade of 2.5 or better were selected as subjects in this study. Their scores in the Classification Battery were used in the discriminant analysis and computation of centour scores.

The results showed that all information concerning the separation of the eight 17-variable centroids could be described by a two-dimensional discriminant space. Centour scores for all groups were computed and used as a basis for the construction of a reference centour table. The size of the latent roots and the overlap of group distributions showed that the Airman Classification Battery achieved a rather small amount of group separation. This indicated that misclassification would have been large for a considerable number of airmen if this battery was used for airmen assignment. The results of the analysis were not validated on check or new samples. The emphasis of this study was to show the use of discriminant analysis and centours for personnel classification and placement problems. Further investigations were still needed in identifiying other variables and other _a priori_ classification of groups which may show a better separation of groups.

Stinson (1958) conducted a discriminant analysis study of engineering students at Oklahoma State University involving four variables and three groups. The four variables used were the total scores on the ACE, the Verbal Comprehension and General Reasoning sub-scores of the Guilford-Zimmerman Aptitude Survey and the Scientific Scale of the Kuder

Preference Record. The three groups differentiated were engineering graduates, non-engineering graduates, and dropouts. Each group consisted of thirty randomly selected subjects who took the test five years before. The purpose of the study was to determine the critical discriminant function scores (v-scores) of the three training categories of students. Using the mean values of the tests of each category, the discriminant function scores were computed three times. The predicted v-score of engineering graduates was 2.048146, the non-engineering graduates was 1.673095, and the dropout was 1.509139. Midway between the predicted v-scores were the values which were considered critical scores. These critical scores were then considered as the basis for predicting the individual's pattern of measured traits. If the predicted v-score of a student was greater than 1.860621, his pattern of measured traits resembled that of the engineering graduate. A v-score between 1.591117 and 1.860621 was taken to indicate that the student was more like those who transferred to another college. A v-score of below 1.591117 probbly indicated the pattern of measured traits of a dropout. This study was intended to show a method for counseling engineering students using the discriminant coefficients as weights in calculating the predicted discriminant function score of an individual. The number of test variables and number of cases studied were quite limited and further validation was needed from a check or fresh sample. However, the study demonstrated the potential uses of the discriminant analysis in answering the question of the student's membership with a training outcome group.

Dunn (1959) investigated the validity between the discriminant and regression methods in a study using 1380 students at Brown University.

Her purpose was to find out whether a student's major field in college was better predicted by discriminant or regression analysis. Fourteen fields of concentration groups and 13 variates consisting mostly of achievement tests were used in this study. An experimental sample of 925 cases who successfully completed the bachelor of arts degree in the 14 fields of concentration were chosen by stratified random selection. Similarly a check sample of 455 cases were selected in order to test the validity of the two techniques.

The discriminant analysis made on the experimental sample showed that the first two discriminant functions accounted for 86 percent of the possible discrimination. These first two discriminants were used in calculating the discriminant weights. Discriminant scores were then computed on the 455 cases in the check sample using the equations established in the analysis of the experimental samples. Chi squares were calculated to obtain the centour scores. Using the same 13 variates and 14 fields of concentration, multiple regression analysis was computed on the experimental samples using the grade point averages in the fields of concentration as the criterion of success.

The results of the calculation of discriminant weights and regression weights showed that variables found as useful predictors in identifying distinct groups in the discriminant analysis were quite different from variates found as useful predictors of success in the regression analysis. The variables which best separated the chemistry field from other fields in discriminant analysis were secondary school rank, mathematics, and science achievement tests. While these three measures were important for success in chemistry, the regression analysis showed that verbal aptitude and reading ability had high weights

instead. These contrasting differences in usefulness of variables as predictors showed in the various fields of concentration in the two analyses.

In order to verify the effectiveness of prediction of the two approaches, two discriminant scores and 14 grade predictions were computed and applied to 455 students in the check sample. For each subject chi squares were computed from his discriminant scores and the three groups which he was most likely a member was determined. Also, the three highest predicted averages computed from the regression weights were selected for each student in the check sample. The predictions by the two methods were tabulated as hits, near hits, and misses. From these results Dunn concluded that the prediction from discriminant scores was far superior than regression scores. Dunn pointed out that using regression scores might lead to a choice of a major field in which the student lacked certain abilities to be successful. Also, it might lead to a choice in which his strongest abilities would not be fully utilized.

Dressel in an appendage to Dunn's article, commented that discriminant and regression procedures offered two different purposes of choices. Regression analysis in this study was made to do something which it was not intended to do. Dressel did not quite agree with Dunn's suggestion to abandon the multiple regression procedure. Dressel explained that it was possible that grade expectation was a normal situation which students encounter in their school life. There were students who entered a field because they were sure of getting high marks and this was usually reasonable with certain conditions. Dressel also raised the social implications involved on this matter and asked whether we were after status quo or upgrading of the person in the field.

He maintained that the multiple regression analysis "holds the possibility of uprading quality of persons if this appears to be desirable." A comparison between these two procedures was explained by Rulon (1951) in a meeting to the American Psychological Association at Pennsylvania State College pertinent portions of which were quoted below:

> The discriminant analysis does what the multiple correlation approach does not. It uses group membership as the criterion and makes all comparisons between groups and none within the groups. ... ... the multiple correlation technique applied separately to several groups and the multiple discriminant technique applied to the several groups simultaneously. ...
>
> Clearly the multiple correlation technique applied to one group ignores all the data from other groups. The multiple discriminant technique employs all the data from all the groups. ... The multiple discriminant technique requires that the same kinds of measurements be made on all the members of all the groups. ...
>
> The multiple correlation technique does not require that the same tests be used in all groups and this is an advantage of a sort, but it means that there is no basis for between-groups comparisons [p. 88].

Tiedeman (1951) pointed out that discriminant analysis is not a replacement of regression analysis. These two types of analysis were useful for two different types of problems. Discriminant analysis was useful in answering the question of a person's resemblance to certain groups while regression analysis helped answer the question of "What am I best at?" Either one of them could be a powerful tool to test a theory.

Calia (1960) investigated the use of discriminant analysis in the prediction of scholastic performance. He was interested in the general problem of academic survival versus attrition. He used the preadmission and first semester data of the entering class of 1957 at Boston University Junior College freshmen as the independent variables for the

prediction of membership in one of three academic groups namely: scholastic failures, terminal prospects, and transfer candidates. While ideally the freshman data of the following year should have been used as cross-validation sample, the discriminant analysis data obtained in this 1957 sample were used to predict the group membership of students from the preceding year's class to obviate waiting for a year before the number of hits and misses could be tallied. Four analyses were carried out in variables assembled in four batteries instead of all 33 variables analyzed simultaneously. Eight variables of the Differential Aptitude Tests constituted the first battery; ten variables from the Kuder Preference Record made up the second battery; six nonintellectual variables from the Survey of Study Habits and Attitude Scores and of two Jervis Self-Description Inventory Scores made up the third battery; and nine variables from high school achievement and activity participation indicies, grade point index, Cooperative English Tests in Vocabulary, Speed, and Level, Otis Gamma Form AM, Scholastic Aptitude Test in Verbal and Mathematics made up the fourth battery. The first three analyses involved only males while the fourth analysis involved both sexes.

The results showed that only one discriminant function was needed to account for membership among groups in each analysis. This indicated the possibility of collinearity with the use of discriminant analysis. On the basis of the Differential Aptitude Tests battery 40 percent were hits on the failure group, 43 percent on the terminal group, and 42 percent on the transfer group. The Kuder Preference Record more or less predicted hits similar to the DAT battery. Among the batteries analyzed the fourth battery, consisting of high school achievement, grade point index, IQ and aptitudes, predicted the best in the study and was followed

by the DAT and Kuder. The non-intellectual battery was not an effective predictor although it was a promising predictor of membership with the DAT and Kuder. Battery 3 and 4 showed as better predictors for the high achievers than for the low achievers. The high relative weight of grade point index as a variable indicated that the nature of one's scholastic achievement was what matters in discriminating among failure, terminal, and transfer groups. For the Boston University Junior College freshmen, academic performance was highly related to scholastic aptitude and for science and humanities, interest represented a factor essential to success.

Tiedeman commented in an appendage to Calia's article that the predominant discrimination of grade point index in this study was not surprising when one analyzed groups in relation to the variable primarily defining them. The result of collinearity of centroids would seem inevitable.

Cass and Tiedeman (1960) conducted a study on vocational development and election of a high school curriculum. They wanted to investigate the reflections of self in vocational development as evidenced through school curriculum choices using variables such as scores of the Otis Quick-Scoring Mental Ability Tests Beta Form CM, the ten sub-scores of the Kuder Preference Record, the Number and Names Tests of the Minnesota Clerical Tests, the Minnesota Paper Form Board, age, sex, family income as estimated from median annual pay in the state of Maine based on the 1950 census. Data were taken from 884 grade nine pupils in nine secondary schools in the state of Maine in the fall of 1951. In the state of Maine, pupils after completing grade 9 may elect any one of the following high school curricula: college, general, commercial, home economics,

industrial arts, and agriculture. After the completion of grade 10 those pupils who took the tests and earned a grade average of 75 or above in the final grades in all subjects were included in the analysis. The number of cases analyzed finally reduced to 466 due to attrition and lack of success. For cross-validation 152 of the cases were selected by stratified random sampling.

The results of the discriminant analysis and Rao's significance test showed that 3 significant roots accounted for 95% of the total discriminating variations of the six groups as measured by 18 variables. The first discriminant function indicated that sex was important in curriculum choice and clearly differentiated boys and girls in the two vocational education areas as interpreted from the weight of this variable and from the relative projections of the group centroids in function 1. Similarly, the second function indicated that age for grade, family income, and academic orientations differentiated the college group from the rest. The third function differentiated interests in outdoor and computational activities among the boys in the two vocational education areas. Curriculum choices were better differentiated by 7 of the Kuder interest scales as shown by the high weights in these variables. The choices were also differentiated by the Minnesota clerical, Bennett mechanical, scholastic aptitude, age, family income, and sex. In this study aptitude showed a minor role in determining curriculum choice.

Generally, the hit rate in the check samples was about 50%. To predict exact curriculum choices only about one-sixth of the cases could be considered hits. To predict college, general, or male or female program of vocational education about one-third of the cases could be

considered hits. The study indicated that in the early high school life of the pupil choices of a curriculum was influenced by sex role, family income, age at that grade, and interests, and less influenced by aptitudes. As pupils entered high school they were little aware of aptitudes. Awareness of aptitudes, as measured by psychological tests, matured late (O'Hara, 1958; Madaus & O'Hara, 1967) and was better comprehended in the senior year of high school. The development of general values seemed to occur throughout the high school period.

One comprehensive study on the prediction of membership of individuals in certain groups was that conducted by King (1958) in predicting choice of undergraduate fields of concentration in Harvard College using multiple discriminant analysis. The 36 predictor variables consisted of sub-scores of the Kuder Preference Record and scores on the academic aptitude survey tests developed in Harvard which included subtests in social science concepts, social science reasoning, arithmetic, biology, diagram interpretation, spatial relations, science, algebra, matrices, camouflaged designs, verbal production, etc. The 36 groups of graduates investigated included such fields as mathematics, physics, engineering, astronomy, chemistry, biochemistry, biology, geology, physical science, psychology, anthropology, social science, economics, government, history, literature, English, classics, romance languages, philosophy, etc. The sample who took the tests consisted of 521 freshmen. A check sample of 299 cases was selected for validation.

The full 36 x 36 among groups and within groups matrices were used in the discriminant analysis using Bryan's method of calculation with some modification. Another discriminant analysis was made on the

10 largest components from the results of the first analysis. Four discriminant roots were found significant in both analyses and were then used to determine the dimensionality of the reduced space. Since the results of both computations were very similar, the rest of the statistical analysis was done on the 10 components instead of the original thirty-six.

Validation with the check samples was tabulated to show direct hits, near misses, and misses. The predicted hits and near misses significantly exceeded chance hits and near misses with a probability well below .001-level. The general conclusion of the study was that interests and aptitudes were effective predictors of undergraduate fields of concentration at Harvard College. However, it would have been interesting to show the percentage of hits than just significance level. The investigator suggested that a similar analysis of a limited number of College Board entrance tests and other admission variables in conjunction with interest or attitude tests self administered after admission but prior to registration could be useful as guidance information to freshmen through their freshmen adviser before selecting a field of concentration after completing the sophomore year.

Stahmann and Wallen (1966) conducted a multiple discriminant prediction of the major field of study at graduation for a sample of students at the University of Utah. The tests used were: the Cooperative English Test Lower Level, Cooperative Mathematics Pre-Test for College Students, Cooperative General Achievement Tests in Natural Science, Occupational Interest Inventory Advanced Level, and the index of urban or rural high school attendance. Random male samples of fifty each from the engineering, business, pharmacy, letters and science fields were selected to constitute the experimental groups upon which the discriminant analyses

were computed. Another fifty random samples from each major field were selected to constitute the cross validation groups upon which the predictions were made. The same procedure was followed in securing the random female samples for nursing, elementary education, and letters and science fields. The discriminant analyses for males and for females were made separately.

The results showed that for the male cross validation, prediction exceeded the number of correct classifications expected by chance for engineering and letters and science but not for the business field. The results of the female cross validation sample was very significant. An analysis was also made between predicted and actual field of study for the male and the female experimental samples using the same discriminant functions of the experimental samples. The results of these predictions for each field of study exceeded chance expectation. However, the results of classification were not shown in percentage of hits which would have been more interesting and meaningful than significant level.

The investigators concluded that the freshmen entrance battery tests at the University of Utah was shown to be an effective predictor of the major field at graduation. They suggested that other variables be used other than achievement and interest measures for the business field group. This study was concerned with academic college fields. Investigations with regard to the prediction in the technical and vocational fields above the high school but below the college grade have seldom been carried out although a far larger number of young adults elected various occupations in this level.

Another similar study on multiple discriminant prediction of college career choices was conducted by Vacchiano and Adrian (1966).

It explored the feasibility of using personality need constructs as measured by the Stern's Activity Index (AI). This measure consisted of 300 need items describing common activities and feelings. The five college groups investigated were business, chemistry, and mathematics male students and education and nursing female students. Like the study of Stahmann and Wallen about 50 students were randomly selected to compose the criterion group and about 35 students to compose each cross validation group.

The results showed only one significant discriminant root which accounted for 72% of the total discriminating variations of the 5 groups. Interpretations were made on the need scales furnishing the greatest relative contribution in the different fields for both the male and female groups based on the discriminant weights in function 1. For example, the business group showed high scores for needs affiliation, aggression, dominance, exhibition, fantasized achievement, humanism, and play. Correct classifications of the male and female groups were above chance expectations. Sixty-five percent of education, 58% of nursing, 54% of business and 33% of mathematics students were correctly predicted.

This study suggested that prediction of academic major might be feasible using personality variables as measured by the Activities Index although other cognitive factors were also important in determining vocational goals.

Summary

Discriminant analysis, which originated with Fisher (1936), was suggested by Travers (1939) and Garret (1943) for application in investigating psychological classification problems. Early investigations using the discriminant analysis in psychology were more concerned

with verifying its effectiveness, in comparing this procedure with regression analysis, and in finding a method of solution for analysis involving more than two groups. The later studies turned to identifying different types of predictor variables. Baggaley (1947) used the sub-scores of the Kuder Preference Record as predictors for two groups of Harvard students. Tiedeman and Steinberg (1951) used the sub-scores of the Differential Aptitude Tests for grade nine pupils. Tiedeman, Bryan, and Rulon (1953) investigated the utility of the Airman Classification Battery as predictor variables in the placement of Air Force personnel. The score on the ACE, some sub-scores of the Guilford-Zimmerman Aptitude Survey and the Scientific scale of the Kuder were used by Stinson (1958) in a discriminant analysis study of engineering students. Other investigators used the achievement tests, nonintellectual variables, intelligence tests, grade point index, high school average, aptitude tests, and personality need constructs. Most of the studies reviewed showed positive results. However, King (1958) pointed out the limitation of the discriminant analysis in producing a direct measure of "satisfaction" or "appropriateness." He cautioned users of discriminant results regarding direct placement of an individual in a field of concentration. The results were better used as a guideline for the student in his election of a course to pursue.

One thing which stood out in the various studies reviewed was that discriminant analysis was a useful statistical tool for psychological studies. So far the various related studies had investigated samples of high school and college students but none of the type of young adult students pursuing the technical and vocational fields above the high

school level but below college grade in institutes of technology. These students constituted a large segment of the future working population who no longer entered the university but deserved some attention with regard to assisting them in choosing suitable occupational training courses. If Super's (1954) view of vocational counseling from the North American cultural framework was accepted as valid, such assistance was surely necessary and important.

If the findings of the studies of O'Hara (1958) and Madaus and O'Hara (1967) that aptitudes were better comprehended by the student in his senior or high school graduation year was accepted, the use of the Differential Aptitude Tests as some of the predictor variables in this study had justification for use in investigating a population of students of the category above. As mentioned earlier, some of these students had outside work experiences. Such work experiences could also contribute to their awareness of the required aptitudes in the real world of work.

CHAPTER IV

METHODS AND MATERIALS

The main interest in this study was to develop a procedure for predicting training outcomes of students enrolled at Northern Alberta Institute of Technology. From a combination of nineteen predictor variables the discriminant and centour classification techniques were used to determine the probability of a new student entering the institute satisfactorily completing his training program on time. The use of this scheme in vocational guidance would assist the incoming student considerably in choosing a program. Students might be counselled to enter a program somewhat different from their initial choice, if their probability of being successful in another program was greater. Students who were likely to achieve a certain unsatisfactory training outcome could be identified beforehand and some preventive steps taken in their guidance and instruction. Furthermore, this procedure might be useful in planning future policies of student selection and placement at NAIT.

Evidence reported in previous studies indicated that the problem of determining, from a combination of predictor variables, which group or occupational category an individual resembles most could best be approached by multiple discriminant analysis and by determining his centour scores. How these approaches could be carried out in a school guidance situation particularly in vocational counselling in technical institutes still needed further study.

The Predictor Variables

The nineteen predictor variables used in this study were eight sub-scores of the Differential Aptitude Tests; two scores of Lorge-Thorndike Intelligence Tests, verbal and nonverbal; five scores of the Alberta Grade IX Departmental Examinations, two scores of the School and College Ability Tests Level 3, verbal and quantitative; and the two high school averages in Grades X and XI. Specifically, the eight sub-scores of the Differential Aptitude Tests were verbal reasoning, numerical ability, abstract reasoning, clerical speed and accuracy, mechanical reasoning, space relations, language usage I-spelling, and language II-grammar. The combination sub-score in verbal reasoning plus numerical ability although available was not included. This kind of combination score would merely produce a linear effect in multivariate analysis and would not in any way contribute to prediction. The five scores of the Alberta Grade IX Departmental Examinations were reading and literature, language, social studies, mathematics, and science. Of the nineteen scores used in this study the first ten of them were gathered after the student had enrolled and chosen his training course. The last nine scores were based on achievement of the student prior to his enrollment at NAIT.

The letter grades in high school subjects were converted by assigning numerical values as follows: H = 5; A = 4; B = 3; C = 2; D = 1. Grades in percent, used in some high school records were also converted as follows: 80% to 100% = 5; 65% to 79% = 4; 50% to 64% = 3; 40% to 49% = 2; 0% to 39% = 1. These particular conversions of grades into these scales was similar to a stanine scale distribution.

The study samples

As stated earlier, the guidance and counseling personnel at NAIT had been concerned with the problem of providing further assistance to students in the matter of choosing a training program. To secure some information about students which would be useful in vocational counseling the Differential Aptitude Tests and the Lorge-Thorndike Intelligence Tests were given to all students who enrolled for the first time in the various technologies, business, and vocational courses about two months after the beginning of the 1966-67 school term. These tests were administered under the supervision of Dr. A. M. Bolle, then head of the Guidance and Counseling Services. Students who took these tests had already chosen and started in their respective courses by the usual procedure prescribed by the institute. It was then assumed that choices made by these students were not in any way influenced by the test results.

The number of students tested and the courses they were enrolled in at that time is shown in Table 1. A total of 1193 students took the Lorge-Thorndike Intelligence Tests. However, not all of them took the Differential Aptitude Tests. This study required that all students must have had a complete set of nineteen scores on the predictor variables. One hundred four students who missed taking the Differential Aptitude Tests were not included in the study sample.

The number of subjects was further reduced by two hundred fifteen due to unavailability of student marks in either Grade X and XI, the Alberta Grade IX Departmental Examinations, or the School and College Ability Tests Level 3. Some students were from other provinces or were former high school graduates who had returned to study after a few

Table 1. Number of students tested by program of study

Technology	D.A.T.	Lorge-Thorndike
Air conditioning & refrig.	8	8
Architectural	53	60
Banking and finance	19	19
Business administration	95	102
Chemical	69	77
Civil	47	54
Commercial cooking	17	17
Dental assisting	28	34
Dental laboratory	15	18
Dietary services	20	21
Distributive	34	53
Drafting	36	47
Electrical	22	24
Electronics	158	166
Exploration	30	30
Forest	55	56
Gas	17	18
Heavy duty mechanics	15	15
Electronic data process	34	40
Industrial production	10	11
Instrumentation	25	25
Material	18	19
Medical laboratory	75	75
Medical X-ray	32	32
Office machine mechanics	15	15
Photography	20	22
Plastics	14	14
Secretarial	56	58
Surveying	11	12
Telecommunications	41	51
T o t a l s	1089	1193

years of absence from school. Scores and marks for these students could not be secured.

Finally, only 868 students were included in this study. The graduates were grouped into nine training fields as follows:

Code	Training field	Specific technology enrolled
1	Control system	Air conditioning & refrig. Electronic data processing Gas Instrumentation
2	Industrial	Heavy duty equipment Industrial production Office machine Plastic
3	Natural resources	Exploration Forest
4	Engineering	Civil Material Surveying
5	Plans and designs	Architectural Drafting Photography
6	Laboratory	Chemical Medical laboratory
7	Health	Dental assisting Dental laboratory Dietary services Medical X-ray
8	Commercial	Banking and finance Business administration Distributive Secretarial
9	Electrical-electronics	Electrical Electronics Telecommunications

In classifying the specific technologies into the various training fields the similarities of aptitudes demanded in the technology and the

characteristics of the working environment were used as bases of consideration. The assumption was that an individual tended to seek occupational classes whose members have a similar personal orientation to his own. The studies of Holland and associates (1966) on vocational choice demonstrated findings which supported this tendency. Some people preferred to seek outdoor rather than indoor work, working among people rather than working alone, structured tasks rather than ambiguous tasks, concrete rather than abstract problems, verbal rather than nonverbal activity, etc. The description of skills and knowledge required of each technology, as described in the institute's calendar, was studied. Comparisons were made among the various technologies to arrive at an appropriate matching of similar aptitude demands and environment orientations of the various training fields. The experiences and opinions of the Director of Instruction of NAIT were also sought in deciding the final classification of training fields.

On the basis of actual training outcomes at the close of the 1967-68 school term, that was two years later, students within the classifications described above were categorized as (a) graduates, i.e. those who graduated in their respective courses on time or successfully completed the second year of their three-year course and (b) non-graduates, i.e. those who failed to meet the graduation requirements, transferred or withdrew before completing a course. Graduation on time or successful completion meant the student passed all subjects with a grade of at least 50%.

The graduates were classified into 9 training fields and a stratified random check sample was drawn and set aside for purposes of

validation. The breakdown of criterion and validation samples classified by training field and technology is shown in Table 2. The number of non-graduates who failed and transferred and withdrew classified by technology is shown in Table 3.

Statistical procedures

The major part of the statistical analysis in this study was carried out on an IBM 360/67 electronic computer which was available at the University of Alberta Computing Center. The investigator sought the assistance of the Division of Educational Research Services of the University in connection with the preparation of the FORTRAN programs for multiple discriminant analysis and centour score classification.

To test the null hypothesis (1) that there are no significant differences in group means in each of the nineteen predictor variables, a one-way analysis of variance was carried out on the criterion sample of graduates.

Five discriminant analyses were done on the sample to determine the most effective combination of training groups and predictors using MULV10 FORTRAN program, (D.E.R.S., 1968). The results derived in each analysis were used in computing the discriminant scores, chi square values and the corresponding conditional and Bayesian probabilities and group classification for each of the graduate criterion sample and validation sample. Comparisons of results were made in terms of the number of predicted correct classifications, close classifications, fairly close classifications, misclassifications, and the corresponding percentages. By close classification was meant that centour score or probability which was next to the correct classification and fairly close

Table 2. Criterion and validation samples of graduates classified by training field and technology

Code No.	Training field	Technology in this classification	Criterion samples	Validation samples	Total
1	Control systems	Electronic data processing Gas Instrumentation	30	7	37
2	Industrial	Heavy duty equip. Indust. production Office machines Plastics	26	6	32
3	Natural resources	Exploration Forest	34	8	42
4	Engineering	Civil Materials Surveying	35	8	43
5	Plans and design	Architectural Drafting Photography	36	9	44
6	Laboratory	Chemical Medical laboratory	54	13	67
7	Health	Dental assisting Dental laboratory Dietary service Medical X-ray	46	11	57
8	Commercial	Banking & finance Business admin. Distributive Secretarial	56	14	70
9	Electrical-electronics	Electrical Electronics Telecommunications	55	14	69
	Total		371	90	461

Table 3. Non-graduates who failed and transferred and withdrew classified by technology

Technology enrolled	Failed and Transferred	Withdrew	Total
Airconditioning and refrigeration	2	3	6
Electronic data processing	5	10	15
Gas	5	3	8
Instrumentation	12	1	13
Heavy duty equipment mechanic	2	1	3
Industrial production	2	2	4
Office machine mechanic	3	0	3
Plastics	1	1	2
Exploration	0	2	2
Forest	6	4	10
Civil	1	8	9
Materials	1	5	6
Surveying	2	2	4
Architectural	17	17	34
Drafting	5	0	5
Photography	0	6	6
Chemical	10	19	29
Medical laboratory	3	12	15
Dental assisting	4	2	6
Dental laboratory	2	1	3
Dietary services	2	2	4
Medical X-ray	5	1	6
Banking and finance	7	2	9
Business administration	22	21	43
Distributive	6	11	17
Secretarial	5	26	31
Electrical	6	7	13
Electronics	46	36	82
Telecommunications	5	11	16
Totals	187	216	403

classification was that centour score or probability which was next to the close classification.

The first analysis and classification involved 9 training groups and 19 predictors. The classification computation in this analysis and in the second analysis were done twice, first by using the common dispersion (MULV11) and the other by using the separate dispersion of each group (MULV12), (D.E.R.S., 1968). These two computations were made in order to verify whether the results of both classification solutions were similar. These comparisons were necessary because too wide differences in group dispersions would affect the effectiveness of classification.

The extent of overlaps between groups in the first analysis was examined. Two of the groups which had a considerable overlap were combined and a second analysis and classification was carried out involving 8 training groups and 19 predictors to find out whether this procedure would improve the effectiveness of classification.

The discriminant weights of the predictors derived in the first analysis were also examined to identify predictors which probably contributed most to differentiation of training groups. Ten predictors with high discriminant weights were selected and used in a third analysis and classification involving the same 9 training groups. The results of this analysis were compared with those in analysis 1 to verify whether 10 predictors would differentiate training groups as effectively as 19 predictors.

To further compare the classification results with those of the first, second, and third analyses, a fourth analysis and classification

was computed involving the 8 training groups and the selected 10 predictors with high weights.

At NAIT the scores of 12 of the predictors in this study were easily available to vocational counselors. Except for Math. 9 (Alberta) and SCAT 3 (quantitative), these readily available 12 predictors consisted of the 10 predictors used in analysis 4 plus DAT LU I (spelling) and Lorge-Thorndike (NV). To determine whether these predictors would be comparatively useful for counseling, a fifth analysis and classification was carried out involving the 8 training groups and these predictors. The results of classification in this analysis were compared with analysis 2 and 4.

Further verification of results in analysis 2 was made by applying the discriminant weights, group centroids, and common dispersion matrices derived from this analysis in computing centours, probabilities, and classifications of 403 non-graduate students. Although it was possible that some non-graduates would be classified as graduates since reasons for non-graduation were not always the lack of aptitude and ability for technical training, the results of this cross-classification would provide additional information of the extent to which this prediction scheme could identify non-graduates from graduates. It would also provide assurance on the applicability of using the results derived in analysis 2 for counseling at NAIT.

Besides the analyses and classifications described above it was felt necessary to investigate the question whether the 19 predictors could differentiate between two broad categories of students, i.e. graduates and non-graduates. As a preliminary step in the counseling process

a counselor would certainly want to know whether an incoming student is likely to graduate from the institute or not before taking up with the student the question of training programs to choose from. Since the counselors and department heads were involved in deciding the admission of new students, information on this question would be valuable in the selection process.

Three hundred twenty-three graduates and 282 non-graduates were used in carrying out a discriminant analysis of two group categories and 19 predictors. The discriminant weights obtained in this analysis were then applied in computing the centours, probabilities, and classifications of 138 validation sample of graduates and 121 non-graduates selected by stratified random sampling.

Multiple discriminant analysis calculation. Bryan (1950) derived the multiple discriminant function matrix equation as:

$$(W^{-1}A - \lambda I)\underline{v} = \underline{0} \tag{1}$$

where W is the within-group sums of squares matrix;

A, the among-group sums of squares matrix;

λ, the latent root;

I, the identity matrix;

and $\underline{v}$, the latent vector. Zero, on the right hand side of equation (1) represents a column vector of zeros. Equation (1) is similar in form to the Kelly or the Hotelling principal axis method of factor analysis and is solved in a similar manner except that $W^{-1}A$ is not symmetric.

The characteristic equation of $W^{-1}A$ is given by the determinantal equation:

$$\left| W^{-1}A - \lambda I \right| = 0 \tag{2}$$

The solution of the latent roots of $W^{-1}A$ in equation (2) satisfies the

discriminant criterion, that is, of providing a maximum value of respective ratios of among-group to within-group sums of squares along the discriminant vectors.

The latent vectors of $W^{-1}A$ are solved by substituting, in equation (1), the value of the latent roots obtained.

These latent vectors represent the corresponding coefficients of the discriminant functions or simply the discriminant weights.

In solving equation (2), Bryan (1950) explained that the number of latent roots λ, which are equal to 0, is at most equal to the number of groups less 1 or equal to the number of variates whichever is smaller. Designating this smaller number as _r_, the total discriminating power of the variates are exhausted by _r_ linear functions defined in this manner. All discriminant functions corresponding to the distinct values of the latent roots λ are uncorrelated as they stand. The general description of the multiple discriminant analysis (Tiedeman, Bryan, and Rulon, 1959) is quoted in full in Appendix I, pp. 142.

The calculation routine of multiple discriminant analysis is as follows:

Let X_{pgi} be the score for person p in group g on test i. Let k be the number of groups, n_g the number of people in group g and m the number of tests. Let N be the total number of people. Define a score matrix X of order N x m of elements x_{pgi} as follows:

$$\underset{(N,m)}{X} = \begin{array}{l} \\ \\ \text{group 1} \\ \\ \\ \\ \text{group 2} \\ \\ \\ \\ \text{group g} \\ \\ \\ \\ \\ \end{array} \overset{\text{tests}}{\left[\begin{array}{cccc} x_{111} & x_{112} & \cdots & x_{11m} \\ x_{211} & x_{212} & \cdots & x_{21m} \\ & x_{p1i} & & \\ & & & x_{n_1 1m} \\ \hdashline x_{121} & x_{122} & \cdots & x_{12m} \\ x_{221} & x_{222} & \cdots & x_{22m} \\ & x_{p2i} & & \\ & & & x_{n_2 2m} \\ \hdashline x_{1g1} & x_{1g2} & \cdots & x_{1gm} \\ x_{2g1} & x_{2g2} & \cdots & x_{2gm} \\ & x_{pgi} & & \\ & & & x_{ngm} \\ \hdashline & & & \\ & & & \\ & & & x_{N_k km} \end{array}\right]} \qquad \begin{array}{l} g = 1, 2, \ldots, k \\ p = 1, 2, \ldots, n_g \ldots, N \\ i = 1, 2, \ldots, m \end{array}$$

1. $\bar{x}_{..i}$ is the grand mean score for test i.

2. The N x m total deviation score matrix for all groups S with elements $x_{pgi} - \bar{x}_{..i}$ has the following form:

$$\underset{(N,m)}{S} = \left[\begin{array}{cccc}
x_{111}-\bar{x}_{..1} & x_{112}-\bar{x}_{..2} & \cdots & x_{11m}-\bar{x}_{..m} \\
x_{211}-\bar{x}_{..1} & x_{212}-\bar{x}_{..2} & \cdots & x_{21m}-\bar{x}_{..m} \\
 & x_{p1i}-\bar{x}_{..i} & & \\
 & & & x_{n_1 1m}-\bar{x}_{..m} \\
\hline
x_{121}-\bar{x}_{..1} & x_{122}-\bar{x}_{..2} & \cdots & x_{12m}-\bar{x}_{..m} \\
x_{221}-\bar{x}_{..1} & x_{222}-\bar{x}_{..2} & \cdots & x_{22m}-\bar{x}_{..m} \\
 & x_{p21}-\bar{x}_{..i} & & \\
 & & & x_{n_2 2m}-\bar{x}_{..m} \\
\hline
 & x_{pgi}-\bar{x}_{..i} & & \\
\hline
 & & & x_{N_k km}-\bar{x}_{..m}
\end{array}\right]$$

3. The m x m total sums of squares score matrix T is the matrix product of

$$\underset{(m,m)}{T} = \underset{(m,N)}{S'} \cdot \underset{(N,m)}{S}$$

4. The mean $\bar{x}_{.gi}$ is the mean score of test i for group g.

5. The N x m within group deviation score matrix S_w with elements $x_{pgi}-\bar{x}_{.gi}$ has the following form:

$$\underset{(N,m)}{S_w} = \left[\begin{array}{cccc}
x_{111} - \bar{x}_{.11} & x_{112} - \bar{x}_{.12} & \cdots & x_{11m} - \bar{x}_{.1m} \\
x_{211} - \bar{x}_{.11} & x_{212} - \bar{x}_{.12} & \cdots & x_{21m} - \bar{x}_{.1m} \\
 & x_{p1i} - \bar{x}_{.1i} & & \\
 & & & x_{n_1 1m} - \bar{x}_{.1m} \\
\hline
x_{121} - \bar{x}_{.21} & x_{122} - \bar{x}_{.22} & \cdots & x_{12m} - \bar{x}_{.2m} \\
x_{221} - \bar{x}_{.21} & x_{222} - \bar{x}_{.22} & \cdots & x_{22m} - \bar{x}_{.2m} \\
 & x_{p2i} - \bar{x}_{.2i} & & \\
 & & & x_{n_2 2m} - \bar{x}_{.2m} \\
\hline
 & x_{pgi} - \bar{x}_{.gi} & & \\
\hline
 & & & \\
 & & & x_{N_k km} - x_{.km}
\end{array}\right]$$

6. The m x m within-group sums of squares matrix W is the matrix product of

$$\underset{(m,m)}{W} = \underset{(m,N)}{S'_w} \cdot \underset{(N,m)}{S_w}$$

7. The m x m among-group sums of squares matrix A is the difference of

$$\underset{(m,m)}{A} = \underset{(m,m)}{T} - \underset{(m,m)}{W}$$

8. The latent roots λ are obtained using the determinantal equation

$$\left| W^{-1}A - \lambda I \right| = \underline{0}$$

9. The latent vectors $\underline{v}$ are computed using the discriminant matrix equation

$$(W^{-1}A - \lambda I)\underline{v} = \underline{0}$$

10. The latent vectors are normalized by multiplying by $1/(\sum_{i=1} \underline{v}_i^2)^{\frac{1}{2}}$; i (variable) = 1, 2, ..., m

The program used also calculated the intercorrelation of the nineteen variables, the standard deviations of the variates in each group and produced punched cards containing the:

1) means of each variate in each group;

2) group dispersion matrices;

3) total dispersion matrix; and

4) transformation matrix (latent vectors of $W^{-1}A$).

The punched cards were later used in the classification solution of the criterion and validation samples.

The program also included the computation of significance tests. A test of significance of the discriminating power of the variates was calculated based on the Wilk's lambda criterion and using the F-ratio approximation developed by Rao (1952).

The test of the dimensionality of the discriminant space was computed using the chi square approximation also developed by Rao (1965). The discriminant functions corresponding to the latent roots found significant at .05-level served as the basis for determining the dimensionality of the reduced test space. The corresponding coefficients of these discriminant functions were then used as the transformation matrix in computing individual discriminant scores.

Classification solutions. Two programs were used in the computation of centours, probabilities, and classifications of the criterion and validation samples in this study. One of these, MULV11 was a solution

using the common dispersion of all groups. The calculation routine of this program is as follows:

Let $\bar{x}_{.gi}$ be the group mean score vector of group g on test i, T the m x m total dispersion matrix, and V_{if} the m x r transformation matrix (latent vectors of $W^{-1}A$) of test i in discriminant function f. The data above are obtained from the results of the discriminant analysis. Let m be the number of tests, r the number functions in the reduced discriminant space, and N the total number of persons.

1. The discriminant score vector $\underline{y}_{pf}$ for person p in function f in the reduced space is

$$\underset{(1,r)}{\underline{y}_{pf}} = \underset{(1,m)}{\underline{x}_{pi}} \cdot \underset{(m,r)}{V_{if}} \qquad \begin{array}{l} p = 1, 2, \ldots, N \\ i = 1, 2, \ldots, m \\ f = 1, 2, \ldots, r \end{array}$$

where $\underline{x}_{pi}$ is the score vector of person p on test i in the test space and V_{if} is the transformation matrix of variable i in function f.

2. The group centroid $\bar{\underline{y}}_{.gf}$ in function f is

$$\underset{(1,r)}{\bar{\underline{y}}_{.gf}} = \underset{(1,m)}{\bar{\underline{x}}_{.gi}} \cdot \underset{(m,r)}{V_{if}} \qquad \begin{array}{l} g = 1, 2, \ldots, k \\ i = 1, 2, \ldots, m \\ f = 1, 2, \ldots, r \end{array}$$

3. The total dispersion matrix D in the reduced space is

$$\underset{(r,r)}{D} = \underset{(r,m)}{V'_{if}} \cdot \underset{(m,m)}{T} \cdot \underset{(m,r)}{V_{if}} \qquad \begin{array}{l} i = 1, 2, \ldots, m \\ f = 1, 2, \ldots, r \end{array}$$

4. The deviation discriminant score vector $\underline{y}_{pg}$ of person p in group g is

$$\underline{y}_{pg} = \left[y_{p1} - \bar{y}_{.g1}, y_{p2} - \bar{y}_{.g2}, \ldots, y_{pf} - \bar{y}_{.gf} \right]$$

$$\begin{array}{l} p = 1, 2, \ldots, N \\ f = 1, 2, \ldots, r \end{array}$$

where $\underline{y}_{pf}$ is the discriminant score vector of person p in function f and $\bar{y}_{.gf}$ is the group centroid vector of group g in function f. As many

deviation discriminant score vectors are computed as there are groups.

5. The position of the discriminant score vector of each person with respect to a particular group in the reduced space is given by the quadratic form

$$\chi^2 = \underset{(1,r)}{\underline{y}_{pg}} \cdot \underset{(r,r)}{D^{-1}} \cdot \underset{(r,1)}{\underline{y}'_{pg}} \qquad \begin{array}{l} p = 1, 2, \ldots, N \\ r = \text{number of functions} \end{array}$$

where $\underline{y}_{pg}$ is the discriminant score deviation vector of person p in each group and D is the total or common dispersion matrix in the reduced space. The quadratic is distributed as chi square with degrees of freedom equal to the dimension of the discriminant space. As many chi square values are computed as there are deviation discriminant score vectors of the person in k groups.

6. The corresponding probability or centour score for each chi square value obtained in step 5 is computed. The centour score may be obtained by multiplying the obtained probability by 100.

7. The conditional classification of a person to a certain group is determined by

$$p(\underline{y}_i \mid H_j) = p(\chi^2 \geq \chi_i^2) \qquad \begin{array}{l} i = 1, 2, \ldots, N \\ j = 1, 2, \ldots, k \end{array}$$

where p is the probability of person i obtaining a score vector $\underline{y}_i$ given that he is a member of the jth group.

8. The probability of group membership is also computed by the Bayesian theorem. First _a priori_ probabilities of i groups are computed as

$$p_j = \frac{n_i}{\sum_{i=1}^{k} n_i}$$

where n_i is the number of persons in group i based on this study. The proposed sizes of enrollment for the new school term may be used for this purpose.

Birnhaum and Maxwell in Cronbach and Gleser (1965) explained the Bayesian probability formula as follows:

> It then follows formally from the formula of Bayes in elementary theory of probability that, if an individual is randomly selected from this hypothetical composite population and if only his measurement u is observed, then the probability that he belongs to category number i is given by $p(i|u) = g_i\, p(u|i)/p(u)$. Here $p(u)$ is the probability that any individual randomly selected from the hypothetical composite population has measurement u, that is,
>
> $p(u) = \overset{k}{\Sigma}\, g_i\, p(u|i)$. ... In this hypothetical context, Bayes' formula suggests the inference that, if an individual is observed to have measurement u, then he is most likely to belong to that category i for which the quantity $g_i\, p(u|i)$ is largest. [pp. 240-241].

g_i is the *a priori* probability that membership in category i is true.

The general equation used in this study was one cited in Cooley and Lohnes (1962) as follows:

$$p(H_j|\underline{y}_i) = \frac{\dfrac{p_j}{|D_j|^{\frac{1}{2}}}\, e^{-\frac{X_j^2}{2}}}{\displaystyle\sum_g \dfrac{p_g}{|D_g|^{\frac{1}{2}}}\, e^{-\frac{X_g^2}{2}}}; \qquad \begin{array}{l} g = 1, 2, \ldots, k \\ i = 1, 2, \ldots, N \\ j = 1, 2, \ldots, g \end{array}$$

where p is the probability that person i, given a discriminant score vector $\underline{y}_i$, is a member of the jth group. This equation is found by substituting in Bayes' probability formula the density function of $\underline{y}_i$ under the multivariate normality assumption. The derivation is shown by Rulon, et al., (1967, pp. 340-342). Bayesian classification of group membership was determined by the highest probability obtained.

In order to describe the extent of overlap between groups in the reduced discriminant space, the chi squares and corresponding probabilities or centours of the centroids in each group were computed using

steps 4, 5, and 6 above. The group centroids in this case were treated as individual discriminant scores.

The MULV12 program was a classification solution using the separate dispersion matrices of each group instead of the common dispersion. This solution was carried out to compare the results of classification with the MULV11 program described earlier. The calculation routine in this program was similar to MULV11 except that the separate group dispersion matrices rather than the total or common dispersion matrix were used in obtaining the chi squares and corresponding probabilities for each person. MULV11 and MULV12 are given in Appendices III & IV, pp. 151, 154.

The concept of centour score. The word centour derives from the combination of "cent" from percent and "our" from contour. The percentage of the area of a normal distribution beyond the contour on which a point lies conveys the idea of centour. (Bryan, 1951) Suppose the mean of a univariate normal distribution is 19 and the standard deviation is 3.72. An individual whose raw score is 24 is equal to a deviation score of +5 and a standard score of +1.344. From a table of areas of normal distribution the proportion which exceeds +1.344 is .089 or 8.9%. In the centour concept a deviation score of +5 is also a deviation of the same magnitude in the opposite direction. For a raw score of 24 the centour score is equal to 17.8. This is the sum of the proportion of the two tails of the distribution as defined by the standard score value. A centour score of 17.8 means that 17.8% of the distribution are more divergent from the centroid (which in this case is the mean of 19) than a raw score of 24, (Rulon, et al.,1967).

In a multivariate distribution for a group, the centour score indicates the percent of score combination which is probably farther

away in all directions from the group centroid. A centour score of 90 indicates that the score combination of the person is near the group centroid while a centour score of less than 1 indicates the score combination of the person is very far away from the group centroid. A centour score can serve as an index for determining whether an individual resembles certain g groups as defined by this test scores, (Rulon, et al., 1967).

For the general case the centour score is determined from the value of chi square which is derived from the exponent of the equation of a normal distribution.

> The locus of points of equal density, sometimes referred to as the locus of equiprobability or iso-frequency, in a normal distribution satisfies the equation
>
> $$x_g \cdot D^{-1} \cdot x_g' = \chi_g^2$$
>
> with x being a row vector of deviation scores for any number of tests T, and with D being the T x T dispersion matrix for the T tests ... In this equation χ_g^2 is an arbitrary constant. The larger values of χ^2 are associated with loci of lesser density. [Rulon, et al, 1967, p. 191].
>
> The fixed value of χ^2 may be determined from the probability table of chi squares. The table gives the value of χ^2 associated with the argument P, the probability of a divergence greater that χ^2 for various degrees of freedom. [Rulon, et al, 1967, p. 118].

The derivation of this quadratic is shown by Rulon, et al. (1967, pp. 115-117). In this study the calculation of the chi squares and corresponding probabilities were included in the computer program instead of referring to a table of chi squares.

For each subject in the validation and criterion samples the computer output the results of the classification solutions and listed

the person's identification number, the vector of discriminant scores, the chi square values and corresponding probabilities of membership in each group, and group numbers indicating the group in which the person was classified. The computer output indicated two types of probabilities for each of the groups: (a) a posteriori probabilities based on Bayes' theorem on a given a priori probabilities and (b) conditional probabilities (centours) based on the chi square distribution for degrees of freedom equal to the dimensionality of the reduced discriminant space. It was an easy matter to convert these conditional probabilities to centour scores by multiplying by 100 or by moving the decimal point two places to the right. However, this was not necessary. Leaving the decimal point as shown in the computer print-out did not in any way alter the concept of centour interpretation.

There were two classification numbers shown. One of these numbers showed the group to which the subject probably belonged if classification was based on the conditional rule (i.e. his highest centour score). The other number was for classification based on the Bayesian rule (i.e. his highest posteriori probability). The output format in these computer classification programs could be used directly by a vocational counselor without any further modification or additional construction of reference tables. (see Chapter VI, p. 130 for explanation)

Since the interest in this study was to develop a procedure for identifying membership of persons in certain training groups for vocational counseling two approaches were considered. After the first run of all the statistical analysis described above the results were studied to find the extent of overlap between training groups. The extent of

overlap could be examined from the centour scores of the group centroids. The larger the centour scores between two group centroids the greater was the overlap of the corresponding ellipsiods and the lesser would be the effectiveness of classification. This was also seen by graphing the discriminant vector scores of the group centroid in the reduced discriminant space.

One of the approaches was to carry out a discriminant analysis and classification solution based on a reduced number of training groups but maintaining the same 19 predictors and find out how well this would improve the classification of groups. The data of two groups which had large overlap were combined as one group in this analysis. Of course, some information was lost in this manner. This also narrowed down the extent of alternatives of classification. However, this drawback could be compensated by an improvement of prediction.

The second approach was to reduce the number of predictors by selecting those predictors with high discriminant weights found in the first analysis and then carry out a third analysis to verify whether differentiation of groups would improve or decrease compared with 19 predictors.

Comparison of results were made in terms of the number of correct classifications, close classifications, fairly close classification, misclassifications, and corresponding percentages. The decision as to which analysis and classification solution was desirable was judged on its utility for vocational counseling.

By studying the positions of group centroids in the reduced space as illustrated in the graphs, psychological interpretations were

speculated upon regarding the meaning of the distribution of scores of the criterion sample.

CHAPTER V

ANALYSIS OF DATA

As shown in Table 2, out of 461 enrollees who graduated a stratified random sampling of 90 cases was set aside for later validation. The remaining 371 cases were used as the criterion sample in the analysis. Four hundred three non-graduate subjects shown in Table 3 were later used in cross-classification to compare the extent of misclassification that would occur if prediction weights of graduates were applied on these students. The first analysis involved 9 training fields and 19 predictor variables.

A one-way analysis of variance was used to test the significance of group difference of means and homogeneity of variance on each of the 19 predictors. The group means of each variable, the F-ratios, the chi square values and the corresponding probabilities are shown in Table 4. All observed F-ratios were significant at .013-level. Hypothesis (1) of no difference in group means was rejected. There were significant differences in group means and this indicated that the variables could be used to discriminate groups in this study. The chi square values were significant at $.91 > p > .06$ except for three predictors, DAT NA, DAT MR, and DAT LU I (spelling) which were significant at $.002 > p > .001$. The data did not contradict the hypothesis of homogeneity of variance except with three variables which perhaps required a larger number of subjects. The main approach of this study was multivariate. Its overall results were quite different from a study of variables taken singly. The

Table 4. Analysis 1: Means, Univariate F-tests, and Chi square tests of homogeneity of variance

	Group means[b]									F^a	p	Chi sq	p
	1	2	3	4	5	6	7	8	9				
1. DAT VR	3.85	3.38	3.77	3.63	3.70	4.10	3.60	3.62	3.63	4.09	.001	11.84	.159
2. " NA	3.61	3.40	3.47	3.58	3.54	3.57	3.33	3.49	3.55	2.71	.006	34.65	.001
3. " AR	4.21	4.01	3.89	3.87	4.17	4.06	3.81	3.94	4.00	2.48	.013	14.92	.061
4. " CSA	5.67	5.53	5.33	5.21	5.33	5.83	5.87	5.42	4.95	5.79	.001	13.74	.089
5. " MR	5.77	5.63	5.53	5.82	5.83	5.15	4.60	5.04	5.80	18.99	.001	25.08	.002
6. " SR	4.32	3.75	3.66	3.96	4.64	3.79	3.43	3.57	4.26	6.06	.001	10.36	.241
7. " LU I	8.74	8.28	8.51	8.00	8.55	8.93	8.92	8.48	7.89	7.17	.001	33.05	.001
8. " LU II	3.92	3.68	3.77	3.30	3.66	4.34	4.16	3.72	3.60	9.19	.001	12.20	.143
9. Lorge-Thrn (v)	6.28	5.65	6.36	6.09	6.25	6.58	6.01	5.99	6.01	3.61	.001	9.32	.316
10. " " (nv)	5.35	5.08	5.18	5.10	5.47	5.20	4.91	4.95	5.27	2.41	.015	7.60	.474
11. Read.9 (Alta.)	6.21	5.42	6.17	5.75	6.06	6.90	6.53	6.07	5.68	7.22	.001	6.50	.592
12. Lang.9 (Alta.)	6.31	5.76	5.94	5.73	5.85	7.09	6.74	6.14	5.73	7.67	.001	5.32	.723
13. So.st.9(Alta.)	6.77	6.03	6.61	6.53	6.47	7.19	6.52	6.48	6.42	3.63	.001	4.07	.850
14. Math.9 (Alta.)	6.69	6.29	6.72	6.64	6.76	7.46	6.26	6.12	6.66	5.70	.001	8.84	.356
15. Sc. 9 (Alta.)	6.95	6.49	6.74	6.89	7.00	7.27	6.45	6.29	6.83	4.01	.001	8.32	.403
16. SCAT 3 (v)	4.36	3.64	4.26	3.93	4.09	4.71	4.37	4.10	4.13	3.52	.001	10.99	.202
17. " " (quant)	4.09	3.72	3.51	3.55	3.88	4.10	3.83	3.78	3.81	2.73	.006	10.87	.209
18. Grade X aver	3.46	3.21	3.43	3.46	3.47	4.01	3.75	3.33	3.42	10.81	.001	3.31	.913
19. Grade XI aver	3.32	3.02	3.29	3.22	3.24	3.86	3.44	3.23	3.11	9.28	.001	5.89	.660

[a] $DF_1 = 8$; $DF_2 = 362$

[b] 1 - Control system; 2 - Industrial; 3 - Nat resources; 4 - Engineering; 5 - Plans & desg; 6 - Laboratory; 7 - Health; 8 - Commerce; 9 - Elec. Electron.

preliminary univariate tests were used at the earlier stage of this study for the purpose of examining the predictive qualities of the variables and the adequacy of sample size. At this stage the preliminary significance tests give favourable results.

Another preliminary examination of the predictive utility of the measures was made by studying the intercorrelations of the 19 variables. The intercorrelations and population means of the criterion sample of graduates are shown in Table 5. As a whole, the intercorrelations were relatively low which was another indication of the favourable possibilities of these variables to differentiate various training groups of graduates. However, out of the 171 pairs of correlations 26 pairs had correlations of .50 to .73. Negative correlations varied from -.002 to -.17 while positive correlations ranged from 0 to .73. Some correlations of .50 or over were found between:

1) DAT AR and DAT SR, Lorge-Thorndike (nonverbal);
2) DAT SR and Lorge-Thorndike (nonverbal);
3) DAT LU II (grammar) and DAT VR, DAT LU I (spelling), Lorge-Thorndike (verbal), Read.-lit. 9 (Alberta), Lang. 9 (Alberta);
4) DAT VR and SCAT 3 (verbal);
5) Lorge-Thorndike (verbal) and DAT VR, Read.-lit 9 (Alberta.), SCAT 3 (verbal);
6) Read.-lit. 9 (Alberta) and DAT VR, Lang. 9 (Alberta) Soc. Stu. 9 (Alberta), SCAT 3 (verbal);
7) Lang. 9 (Alberta) and DAT LU I (spelling), Soc. Stu. 9 (Alberta), SCAT 3 (verbal), Grade X average, Grade XI average;

Table 5. Analysis 1: Intercorrelation and criterion means

	Predictor variables																			Criterion means
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	
1	1.0																			3.71
2	.33	1.0																		3.50
3	.45	.42	1.0																	3.99
4	.09	.10	.08	1.0																5.46
5	.18	.20	.35	-.12	1.0															5.41
6	.30	.32	.55	.10	.61	1.0														3.91
7	.39	.18	.15	.24	-.12	-.04	1.0													8.48
8	.55	.29	.27	.24	.00	.14	.52	1.0												3.82
9	.73	.31	.40	.03	.23	.30	.40	.52	1.0											6.15
10	.32	.41	.55	.22	.33	.50	.22	.26	.35	1.0										5.16
11	.64	.26	.31	.16	-.03	.09	.44	.57	.66	.18	1.0									6.13
12	.50	.28	.20	.22	-.14	.01	.53	.59	.48	.28	.65	1.0								6.19
13	.41	.28	.18	.11	.03	.06	.28	.34	.47	.17	.53	.52	1.0							6.59
14	.34	.45	.32	.13	.12	.21	.26	.31	.36	.37	.40	.49	.47	1.0						6.64
15	.37	.34	.23	.01	.23	.20	.17	.29	.43	.20	.43	.41	.62	.62	1.0					6.77
16	.61	.22	.23	.04	.07	.11	.38	.48	.67	.16	.67	.52	.44	.27	.37	1.0				4.21
17	.37	.39	.36	.16	.09	.16	.29	.30	.40	.31	.41	.43	.35	.51	.40	.55	1.0			3.82
18	.27	.26	.14	.23	-.17	-.002	.33	.41	.27	.14	.41	.56	.50	.49	.50	.25	.30	1.0		3.53
19	.32	.23	.12	.19	-.14	-.002	.33	.40	.34	.11	.41	.53	.45	.39	.38	.28	.27	.72	1.0	3.33

Legend:

Predictor 1 - DAT VR
2 - " NA
3 - " AR
4 - " CSA
5 - " MR
6 - " SR
7 - " LU I (spell.)
8 - " LU II (gram.)
9 - Lorge-Thorndike (V)
10 - Lorge-Thorndike (NV)
11 - Read.-lit. 9 (Alberta)
12 - Lang. 9 (Alberta)
13 - Soc. stud. 9 (Alberta)
14 - Math. 9 (Alberta)
15 - Science 9 (Alberta)
16 - SCAT 3 (verbal)
17 - SCAT 3 (nonverbal)
18 - Grade X average
19 - Grade XI average

8) Science 9 (Alberta) and Soc. Stu. 9 (Alberta), Math. 9 (Alberta), Grade X average;

9) SCAT 3 (verbal) and SCAT 3 (quantitative);

10) Grade X and Grade XI average.

Considering the similarities of the characteristics measured by these variables the correlations of .50 or over were quite possible. For example, DAT VR and Lorge-Thorndike (verbal) which were quite similar measures had a correlation of .73. Between Math. 9 (Alberta), and Science 9 (Alberta) the correlation was .62. DAT NA and DAT CSA showed consistently low correlations with all the other variables.

This study hypothesized that the scores made by individuals on the nineteen predictor variables were related to their membership in occupational groups, that is, these variables could be used to determine different regions in an $\underline{m}$ dimensional test space which certain groups tended to occupy. A preliminary verification of this hypothesis was made by testing the significance of the difference in group centroids by using the F-ratio approximation developed by Rao (1952). The results of these tests are shown at the bottom of Table 6. The lambda obtained was .263 and the corresponding F-ratio of 3.33 was significant at the .001-level. It was concluded, therefore, that the group centroids in the graduate criterion sample were not the same. The generalized multivariate null hypothesis (2) that the characteristics of the various training groups, as measured by the nineteen predictor variables, were the same was not tenable. The approximation test of significance of the discriminating power of the variables further indicated that they could possibly be effective predictors for identifying membership in certain training groups of students at NAIT.

Analysis 1 (9 training groups, 19 predictors). The roots of $W^{-1}A$ and the chi square tests of significance of dimensionality of the discriminant space are shown in Table 6. The maximum of 8 linear roots were obtained in this analysis. The first three roots were significant at .003-level. From these results the first three discriminant functions, whose roots were highly significant, were used as the new dimension of the reduced space. The original 19-dimensional test space was then reduced to a 3-dimensional discriminant space. The corresponding discriminant weights, as shown in Table 7, were used as the transformation matrix in converting the orginal scores to discriminant scores for each subject of the criterion and validation samples.

The standard deviations of the variables for each of the nine training groups are shown in Table 7. Variations in standard deviations among the nine training groups roughly explained the high or low weighting of a particular variable. For example, Soc. Stu. 9 (Alberta) consistently showed low weights in all the three discriminant functions. Standard deveiations of this variable among the nine training groups also showed very little variations in values.

The standard deviations of DAT MR varied from .4 to .9. Its weight in discriminant function I was high. The Grade XI average had low weights in discriminant function I and III but somewhat high weights in function II. Standard deviations of this variable were similar in some training groups but varied by .2 in other groups. The variations in standard deviations, in a general way, could explain the high or low discriminant weighting on a particular variable.

Table 6. Analysis 1: Roots, Chi square test of dimensionality of discriminant space, and multivariate F-test

Root	λ	DF	Chi square	p	Percent
1	.9230	152	464.17	.001	55.59
2	.2998	126	236.61	.001	18.05
3	.1334	102	145.35	.003	8.03
4	.1139	80	101.77	.051	6.86
5	.0763	60	64.23	.331	4.59
6	.0658	42	38.65	.619	3.96
7	.0365	26	16.48	.924	2.19
8	.0121	12	4.01	.983	.73
				Total	100.00

Lambda = .263
DF_1 = 152; DF_2 = 2565
F = 3.33
p <.001

Table 7. Analysis 1: Discriminant weights and standard deviations of groups

Predictor Variables	Discriminant Function I (Normalized Weights)	II	III	Standard deviation Control syst.	Indus-trial	Nat. res.	Engin-eering	Plans & design	Labor-atory	Health	Commer-cial	Elect'l. electron.
1 DAT VR	-.133	-.342	-.120	.495	.717	.569	.548	.655	.525	.713	.675	.692
2 " NA	-.538	.015	-.613	.260	.435	.371	.289	.330	.289	.503	.393	.275
3 " AR	.140	.305	.129	.408	.466	.636	.493	.452	.509	.644	.524	.447
4 " CSA	.154	-.048	-.068	1.10	.869	1.05	.708	.891	.869	.891	.815	.663
5 " MR	-.612	-.181	-.015	.711	.584	.691	.439	.606	.574	.864	.777	.572
6 " SR	.113	.121	.377	.906	.127	1.13	1.01	.796	.926	1.02	.115	1.00
7 " LU I	.092	.114	.132	.649	.929	.803	.937	.796	.733	.806	1.31	1.03
8 " LU II	.389	.080	.283	.549	.685	.661	.577	.713	.751	.838	.580	.684
9 Lorge-Thorndike (V)	-.077	-.141	-.279	.946	.952	.867	.624	.944	.809	.868	.947	.949
10 " " (NV)	-.148	-.042	.080	.763	.866	.922	.800	.661	.706	.671	.728	.700
11 Read. 9 (Alta.)	.141	.002	-.016	.991	1.14	.933	.936	1.24	1.12	1.25	1.08	1.11
12 Lang. 9 (Alta.)	.033	.165	-.064	.115	1.16	1.12	1.06	1.25	1.29	1.40	1.15	1.12
13 Soc.stud. 9 (Alta.)	.012	.029	-.096	1.05	1.24	1.07	1.04	1.09	.986	1.13	1.01	.937
14 Math. 9 (Alta.)	-.080	-.350	-.036	1.44	1.25	1.06	1.15	1.17	1.02	1.07	1.06	1.33
15 Sc. 9 (Alta.)	-.063	.095	.205	1.04	1.10	1.18	.824	1.07	.923	1.19	1.14	1.10
16 SCAT 3 (V)	-.015	-.187	-.114	1.16	.927	.875	.760	1.18	1.01	.890	.971	1.09
17 SCAT 3 (quant)	.147	.362	.362	.891	.816	.620	.769	.785	.712	.595	.851	.777
18 Grade X aver.	.120	-.551	.273	.491	.496	.526	.476	.441	.491	.557	.461	.469
19 Grade XI aver.	.074	-.265	-.047	.598	.511	.619	.447	.569	.516	.570	.481	.577

The variables with high weights in discriminant function I were DAT LU II (grammar), DAT MR, DAT NA while those with very low weights were Soc. Stu. 9 (Alberta), SCAT 3 (verbal), and Lang. 9 (Alberta).

In discriminant function II high weights were found in SCAT 3 (quantitative), DAT AR, Grade X average, Math. 9 (Alberta), and DAT VR while those containing very low weights were Read.-lit. 9 (Alberta), DAT NA, and Soc. Stu. 9 (Alberta).

In discriminant function III, DAT SR, SCAT 3 (quantitative), DAT NA, and Lorge-Thorndike (verbal) were weighted high, while SCAT 3 (verbal), DAT MR, and Read.-lit. 9 (Alberta) were weighted low.

The group mean vectors in the reduced discriminant space are shown in Table 8. The nine group centroids in the reduced space are graphically illustrated in Figure 2. From this figure, centroids 1 (control system), 2 (industrial), 5 (plans and design), and 9 (electrical-electronics) were quite close to each other. The ellipsoids of these three groups overlapped to some extent. Correct classification of persons in these groups would not be as efficient as those in groups 3, 4, 6, 7, and 8, which were quite apart.

To further examine the extent of overlap, the centroid centours for each training group are shown in Table 9. Centour in this table is based on the assumption that there is constant dispersion within each group. For example, if the centroid of group 1 falls on the X percentile ellipse of the multivariate normal distribution then we shall say that group 1 centroid is an X centour with respect to group 2. The pairs of centours of training groups 1, 2, 5, and 9 taken two at a time were quite high. This meant that their centroids were close to each other and indicated a high degree of overlap of their isofrequency ellipsoids.

Table 8. Analysis 1: Group centroid vectors in discriminant space

Group No.	Training group	Discriminant function I	II	III
1	Control system	-1.60	-3.13	2.37
2	Industrial	-1.67	-2.84	2.22
3	Natural resources	-1.69	-3.52	1.87
4	Engineering	-2.16	-3.46	1.90
5	Plans and design	-1.81	-3.20	2.49
6	Laboratory	-0.96	-3.80	2.23
7	Health	-0.50	-3.09	2.17
8	Commercial	-1.25	-2.94	1.90
9	Electric-electronics	-2.02	-3.26	2.26

Table 9. Analysis 1: Centours of group centroids in discriminant space

Group Centroid	Training groups 1	2	3	4	5	6	7	8	9
1	100								
2	95	100							
3	72	58	100						
4	54	48	86	100					
5	97	86	65	62	100				
6	40	19	48	14	28	100			
7	25	19	13	02	11	49	100		
8	75	82	61	29	49	37	54	100	
9	88	80	79	89	95	19	05	44	100

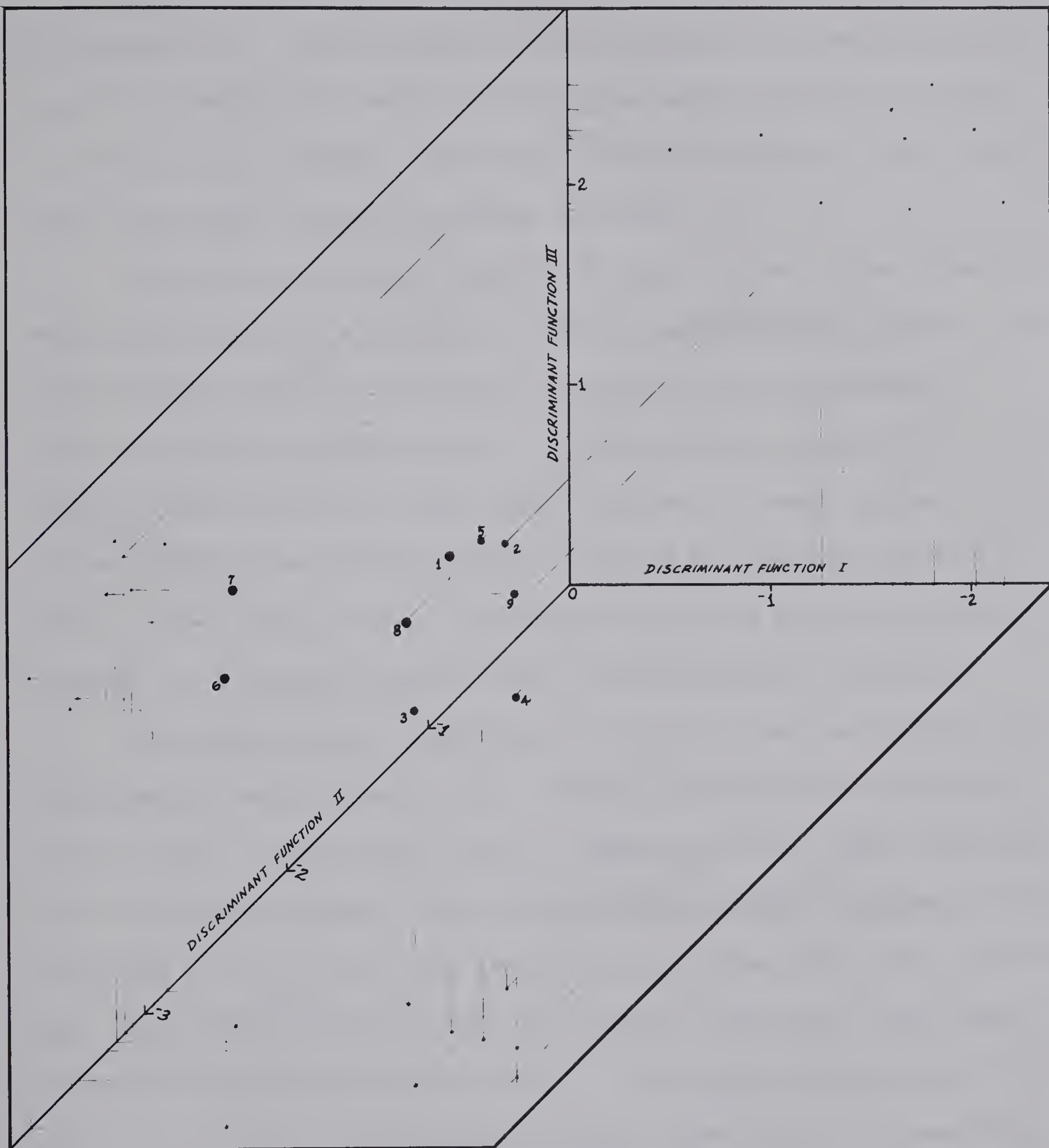

Figure 2. Analysis 1: Location of group centroids in discriminant space

Correct classifications of persons in these groups were expected to be few. Predicted classifications were determined of the subjects in the criterion and validation samples. Two kinds of classification solutions were carried out. One of these solutions computed the centour scores, Bayesian probabilities, and classifications using the total or common dispersion of all groups. The results of classification of the criterion sample using this solution are shown in Table 10.

Under the conditional rule 34% of the subjects in the criterion sample were correctly classified. If the counselor would consider the second highest centour as a close classification the percentage of prediction would increase to 58%. If the counselor considered the third highest centour as fairly close classification, it would further increase the prediction to an overall result of 71% of the subjects in the criterion sample. These results indicated that the 19 predictors differentiated the 9 training groups in the criterion sample fairly well.

Under the Bayesian classification rule for the same subjects in the criterion sample, there was no correct classification in training groups 1 and 2. There were, however, close and fairly close classification in these two groups. The number of correct classifications in training groups 6, 7, 8, and 9 were far higher than those under the conditional rule. The Bayesian rule did well when sizes of the groups were large but not so when group sizes were small. Under this classification rule, 37% of the subjects in the criterion sample were correctly classified. If the results of the close and fairly close classification were included, then 70% might be considered as correctly predicted. This result was about the same as the conditional classification but again was higher

Table 10. Analysis 1: Classification of criterion sample using common dispersion

Training Group	No. in Sample	Type of Classification: Conditional rule: correct	close	fairly close	Total	Miss	Bayesian rule: correct	close	fairly close	Total	Miss
1 Control system	30	3 (2.4)	8	7	18	12	0 (2.4)	4	7	11	19
2 Industrial	26	8 (1.8)	3	5	16	10	0 (1.8)	8	2	10	16
3 Natural resources	34	10 (3.1)	7	7	24	10	4 (3.1)	13	5	22	12
4 Engineering	35	18 (3.3)	7	3	28	7	11 (3.3)	13	4	28	7
5 Plans and design	35	11 (3.3)	7	5	23	12	9 (3.3)	5	5	19	16
6 Laboratory	54	25 (7.9)	12	6	43	11	34 (7.9)	8	2	44	10
7 Health	46	29 (5.7)	9	1	39	7	29 (5.7)	10	1	40	6
8 Commercial	56	10 (8.5)	20	6	36	20	19 (8.5)	17	4	40	16
9 Electrical-electronics	55	11 (8.2)	17	9	37	18	30 (8.2)	12	4	46	9
Total	371	125	90	49	264	107	136	90	34	260	111
Percent	100	34	24	13	71	29	37	24	9	70	30

Note: Numbers in parenthesis are number of classifications based on random assignment of subjects to groups in proportion to the total number of subjects in each group.

than those in the validation samples shown in Table 13. The small number of classifications in training groups 1 and 2 is likely due to the large overlap of these groups as shown in the graphical plot of the group centroids in Figure 2 and due to the small sample size in these groups.

Centours, Bayesian probabilities, and classification were also computed on the same subjects in the criterion sample by the second solution, that is, using the separate dispersion of each group instead of the common dispersion. Predicted correct classifications under the conditional rule as shown in Table 11 indicated an increase in groups 2, 3, 6, 8, and 9 but a decrease in groups 5 and 7. There was an increase in the number of predicted correct classification from 34% to 37%

Under the Bayesian rule, groups 1 and 2 were not well predicted. There were some increases in the number of correct classifications in groups 2, 3, 4, 6, and 8 but they were cancelled out by the decrease in correct classifications in groups 5, 7, and 9. The predicted correct classification was 37% of the subjects in the criterion sample. As a whole, there was slight improvement in prediction when the separate dispersion of each group, instead of the common dispersion, was used in computing the centours, probabilities, and classification. For this sample it indicated that the results in both solutions were not markedly different.

Similarly, the two classification solutions were applied to the subjects in the validation sample. The results of predicted classification, using the common dispersion is shown in Table 12. None of the

Table 11. Analysis 1: Classification of criterion sample using the separate dispersion

Training Groups	No. in Sample	Type of Classification: Conditional rule: correct	close	fairly close	Total	Miss	Bayesian rule: correct	close	fairly close	Total	Miss
1 Control system	30	3 (2.4)	9	7	19	11	0 (2.4)	7	10	17	13
2 Industrial	26	10 (1.8)	5	3	18	8	1 (1.8)	7	6	14	12
3 Natural resources	34	13 (3.1)	6	7	26	8	6 (3.1)	11	6	23	11
4 Engineering	35	13 (3.3)	7	3	23	12	17 (3.3)	9	2	28	7
5 Plans and design	35	7 (3.3)	9	3	19	16	6 (3.3)	11	6	23	12
6 Laboratory	54	29 (7.9)	14	2	45	9	37 (7.9)	9	0	46	8
7 Health	46	28 (5.7)	11	1	40	6	26 (5.7)	14	1	41	5
8 Commercial	56	17 (8.5)	20	6	43	13	20 (8.5)	15	6	41	15
9 Electrical-electronics	55	17 (8.2)	13	7	37	18	26 (8.2)	14	6	46	9
Total	371	137	94	39	270	101	139	97	43	279	92
Percent	100	37	25	11	73	27	37	26	12	75	25

Note: Numbers in parenthesis are number of classifications based on random assignment of subjects to groups in proportion to the total number of subjects in each group.

Table 12. Analysis 1: Classification of validation sample using the common dispersion

Training Group	No. in Sample	Conditional rule correct	Conditional rule close	Conditional rule fairly close	Conditional rule Total	Conditional rule Miss	Bayesian rule correct	Bayesian rule close	Bayesian rule fairly close	Bayesian rule Total	Bayesian rule Miss
1 Control system	7	0 (.55)	0	1	1	6	0 (.55)	0	1	1	6
2 Industrial	6	2 (.40)	0	1	3	3	0 (.40)	2	1	3	3
3 Natural resources	8	0 (.71)	2	0	2	6	0 (.71)	2	0	2	6
4 Engineering	8	5 (.71)	1	1	7	1	4 (.71)	1	2	7	1
5 Plans and design	9	3 (.90)	3	1	7	2	1 (.90)	5	1	7	2
6 Laboratory	13	3 (1.9)	3	2	8	5	3 (1.9)	4	1	8	5
7 Health	11	6 (1.3)	3	0	9	2	6 (1.3)	3	0	9	2
8 Commercial	14	3 (2.2)	2	1	6	8	3 (2.2)	2	4	9	5
9 Electrical-electronics	14	2 (2.2)	5	3	10	4	8 (2.2)	1	2	11	3
Total	90	24	19	10	53	37	25	20	12	57	33
Percent	100	27	21	11	59	41	28	22	13	63	37

Note: Numbers in parenthesis are number of classifications based on random assignment of subjects to groups in proportion to the total number of subjects in each group.

individuals in groups 1 and 3 under the conditional rule and groups 1, 2, and 3 under the Bayesian rule were correctly classified. This was likely due to the limited size of check samples in these groups and to the large overlap between these groups. A total of 27% to 28% was correctly classified under both rules. Considering that the probability of predicting the correct group to which a person belonged was 1 in 9 or 0.11 the results were fairly good. The overall predicted classification under both rules, close and fairly close included, was 57% to 63%.

The classification of the subjects in the validation sample using the separate dispersion of each group is shown in Table 13. The pattern of results was not so different compared with the results using the common dispersion. Under both conditional and Bayesian rules the results accounted for 22% to 26% of the validation samples correctly predicted. The overall predicted classifications were 60% to 62% correct. In these data the results of both classification solutions were not markedly different.

In the two classification rules, it was noted that when the person was correctly classified by the conditional rule but misclassified by the Bayesian rule, the probability under the Bayesian rule was often a close one, that is, if one considered a close or fairly close classification as a correct classification then both the conditional and the Bayesian rules were performing similarly in classifying the person. Also, when the person was correctly classified by the Bayesian rule but misclassified by the conditional rule, the conditional centour value was often close to the highest centour. In other words, if the counselor considered the classifications given in both the conditional and Bayesian rules, he

Table 13. Analysis 1: Classification of validation sample using the separate dispersion

Training Groups	No. in Sample	Conditional rule: correct	Conditional rule: close	Conditional rule: fairly close	Conditional rule: Total	Conditional rule: Miss	Bayesian rule: correct	Bayesian rule: close	Bayesian rule: fairly close	Bayesian rule: Total	Bayesian rule: Miss
1 Control system	7	0 (.55)	1	0	1	6	0 (.55)	1	0	1	6
2 Industrial	6	1 (.40)	1	2	4	2	0 (.40)	1	1	2	4
3 Natural resources	8	0 (.71)	1	3	4	4	0 (.71)	1	1	2	6
4 Engineering	8	5 (.71)	1	0	6	2	5 (.71)	1	0	6	2
5 Plans and design	9	0 (.90)	3	3	6	3	1 (.90)	3	3	7	2
6 Laboratory	13	4 (1.9)	2	3	9	4	4 (1.9)	2	3	9	4
7 Health	11	6 (1.3)	3	1	10	1	5 (1.3)	3	0	9	2
8 Commercial	14	1 (2.2)	4	3	8	6	2 (2.2)	3	2	7	7
9 Electrical-electronics	14	3 (2.2)	3	2	8	6	6 (2.2)	4	1	11	3
Total	90	20	19	17	56	34	23	20	11	54	36
Percent	100	22	21	19	62	38	26	22	12	60	40

Note: Numbers in parenthesis are number of classifications based on random assignment of subjects to groups in proportion to the total number of subjects in each group.

could not be very far from naming the training group to which a person probably belonged, providing the scheme was predicting sufficiently beyond probability expectation. The classification results showed that the nineteen variables could discriminate different training groups of students at NAIT.

Analysis 2 (8 training groups, 19 predictors). As illustrated in Figure 2, the centroids of groups 1, 2, and 5 were quite close to each other. Furthermore, as shown in Table 9 these groups overlapped considerably. The approach was to do a discriminant analysis and classification solution based on 8 training groups, that is, combining groups 1 and 2 and leaving the remaining groups as they were. The nature of the training and aptitude requirement of groups 1 (control system) and 2 (industrial), which consisted mostly of mechanical skills with tools and instruments were somewhat similar and would not appreciably change the picture of the *a priori* groupings. This procedure would probably improve the number and accuracy of correct classification. The breakdown of criterion and validation samples in this analysis is shown in Table 14. Groups 1 and 2 combined consisted of 55 graduates. In this second analysis the first three roots were found to be significant at the .002-level, as shown in Table 15. Three discriminant functions were used to determine the dimensionality of the reduced space. The F-ratio approximation, for testing the difference in group centroids, was 3.56. This was significant at the .001-level.

Discriminant weights of the 19 variables are shown in Table 16. Twelve variables showed high discriminant weights. They were the DAT VR, DAT NA, DAT AR, DAT CSA, DAT MR, DAT SR, DAT LU I, DAT LU II, Math. 9

Table 14. Analysis 2: Criterion and validation samples of graduates classified by training fields and technology

Code No.	Training field	Technology in this classification	Criterion sample	Validation sample	Total
12	Control sys-industrial	Electronic data processing Gas Instrumentation Heavy duty equip. Indust. production Office machines Plastics	55	14	69
3	Natural resources	Forest Exploration	34	8	42
4	Engineering	Civil Materials Surveying	35	8	43
5	Plans and design	Drafting Architectural Photography	35	9	44
6	Laboratory	Chemical Medical	54	13	67
7	Health	Dental assisting Dental laboratory Dietary Medical X-ray	46	11	57
8	Commercial	Banking & finance Business admin. Distributive Secretarial	56	14	70
9	Electrical-electronics	Electrical Electronics Telecommun.	55	14	69
	Total		370	91	461

Table 15. Analysis 2: Roots, Chi square test of dimensionality of discriminant space, and multivariate F-test

Root	λ	DF	Chi square	p	Percent
1	.9252	133	434.14	< .001	59.29
2	.2552	108	205.86	< .001	16.35
3	.1357	85	126.64	.002	8.69
4	.0959	64	82.30	.061	6.14
5	.0644	45	50.40	.268	4.12
6	.0519	28	28.66	.430	3.32
7	.0327	13	11.02	.609	2.09
				Total	100.00

Lambda = .288

DF_1 = 133; DF_2 = 2284

F = 3.56

p < .001

Table 16. Analysis 2: Discriminant weights of three functions

Predictor variables	Discriminant function I	II	III
	(normalized weights)		
1. DAT VR	-.047	-.226	-.113
2. " NA	-.522	-.009	-.249
3. " AR	.158	.257	-.142
4. " CSA	.174	-.055	-.202
5. " MR	-.647	-.195	-.177
6. " SR	.123	.199	.291
7. " LU I (spell)	.120	.239	-.211
8. " LU II (grammar)	.354	-.025	.233
9. Lorge-Thorndike (V)	-.097	-.162	-.158
10. " " (NV)	-.157	-.066	.108
11. Read.-lit. 9 (Alberta)	.101	-.013	-.002
12. Lang. 9 (Alberta)	.015	.143	.022
13. Soc. Stud. 9 (Alberta)	.001	.062	-.077
14. Math. 9 (Alberta)	-.053	-.374	-.060
15. Science 9 (Alberta)	-.048	.055	.054
16. SCAT 3 (V)	.001	-.170	.071
17. SCAT 3 (quantitative)	.126	.391	.259
18. Grade X average	.125	-.562	.627
19. Grade XI average	.124	-.221	-.366

(Alberta), SCAT 3 (quantitative), Grade X average, and Grade XI average. Except for the addition of DAT CSA, DAT LU I, Grade XI average, and the exclusion of Lorge-Thorndike (V), these variables were the same as those found with high weights in analysis 1. Similar observations were also noted in a subsequent analysis 3.

From the centours of the group centroids in this analysis, as shown in Table 18, the large overlaps were in groups 1, 2, 3, 4, 5, and 9. The relative location of the group centroids in the reduced space illustrated in Figure 3, showed that groups 3 and 5 were very close to each other if viewed on the plane bounded by function II and III. The clustering of group centroids was reduced slightly.

Results of classification of subjects in the criterion and validation samples are shown in Tables 19, 20, 21, and 22. Compared with the results in analysis 1 the predicted classifications of subjects in the validation sample increased from 27% to 33% correct. The overall predicted classifications increased from 60% to 71% correct. However, groups 3 and 5 were not well classified.

The classification results of subjects in the criterion sample were also higher than in analysis 1. This improvement in prediction in this second analysis was likely due to the reduction of overlap between groups when training groups 1 and 2 were combined into one group.

The use of either the common dispersion or the separate dispersion of each group in computing the classifications gave similar results. In the later classification computation done in this study, the common dispersion instead of the separate dispersion of each group was used to minimize calculations. It was demonstrated in the second analysis that the

Table 17. Analysis 2: Group centroid vectors in discriminant space

Group No.	Training Group	Discriminant function I	II	III
12	Control sys- industrial	-1.32	-2.44	-2.32
3	Natural resources	-1.33	-2.92	-2.46
4	Engineering	-1.79	-2.88	-2.34
5	Plans and design	-1.44	-2.55	-2.13
6	Laboratory	-0.55	-3.19	-2.24
7	Health	-0.14	-2.48	-2.11
8	Commercial	-0.90	-2.34	-2.40
9	Electrical-electronics	-1.67	-2.69	-1.98

Table 18. Analysis 2: Centours of group centroids in discriminant space

Group Centroid	Training group 12	3	4	5	6	7	8	9
12	100							
3	84	100						
4	73	85	100					
5	96	77	80	100				
6	29	49	15	27	100			
7	19	12	02	14	53	100		
8	89	64	32	69	42	50	100	
9	74	60	83	95	16	05	35	100

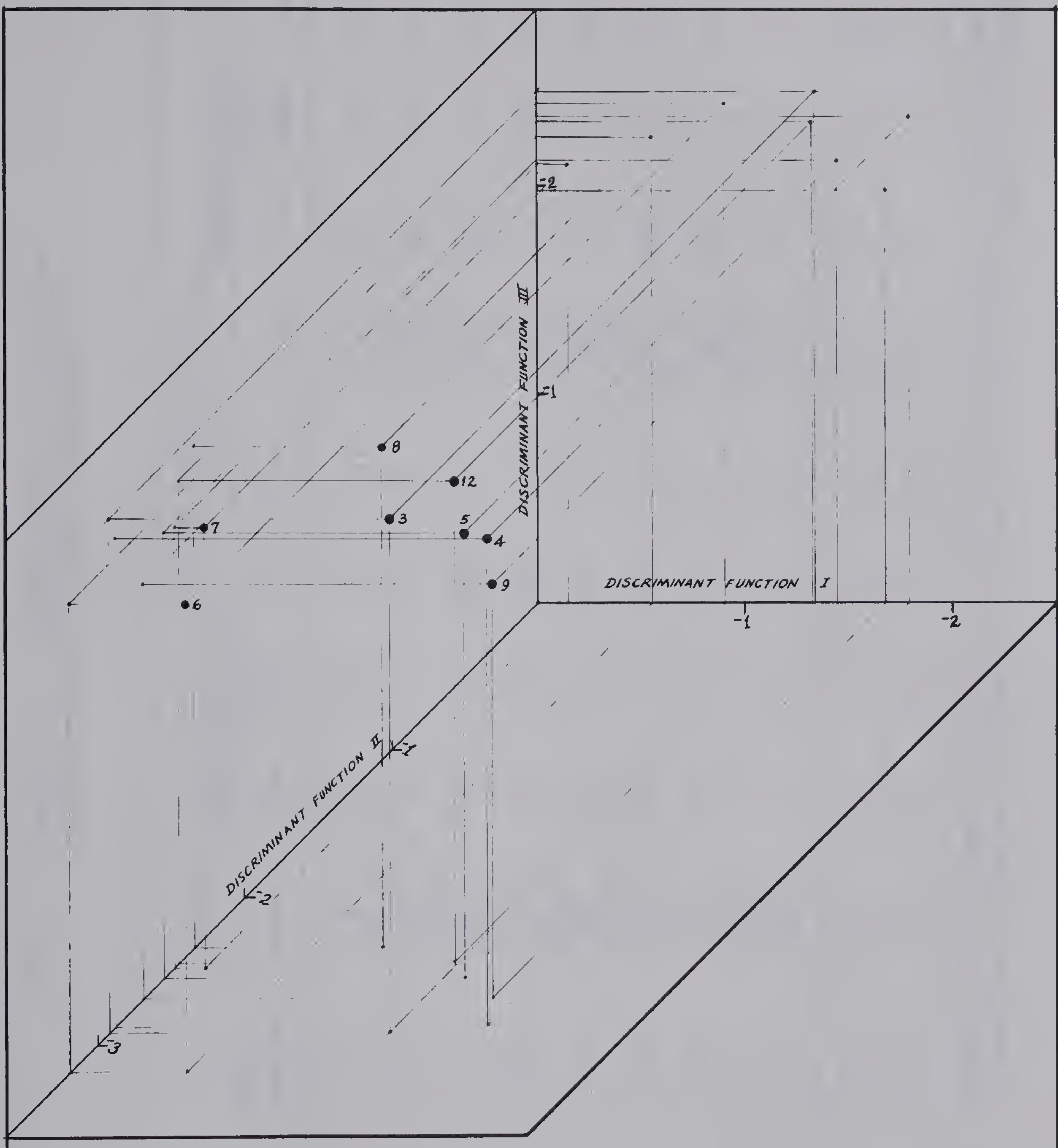

Figure 3. Analysis 2: Location of group centroids in discriminant space

Table 19. Analysis 2: Classification of criterion sample using the common dispersion

Training Groups	No. in Sample	Conditional rule: correct	Conditional rule: close	Conditional rule: fairly close	Conditional rule: Total	Conditional rule: Miss	Bayesian rule: correct	Bayesian rule: close	Bayesian rule: fairly close	Bayesian rule: Total	Bayesian rule: Miss
1,2 Control sys., Industrial	55	9 (8.2)	17	12	38	17	13 (8.2)	24	9	46	9
3 Natural resources	34	12 (3.1)	7	5	24	10	3 (3.1)	13	7	23	11
4 Engineering	35	15 (3.3)	8	5	28	7	11 (3.3)	10	5	26	9
5 Plans and design	35	3 (3.3)	16	7	26	9	0 (3.3)	8	14	22	13
6 Laboratory	54	29 (7.9)	11	4	44	10	33 (7.9)	6	6	45	9
7 Health	46	27 (5.7)	11	3	41	5	27 (5.7)	12	2	41	5
8 Commercial	56	15 (8.5)	17	3	35	21	18 (8.5)	19	3	40	16
9 Electrical-electronics	55	29 (8.2)	6	4	39	16	34 (8.2)	4	4	42	13
Total	370	139	93	43	275	95	139	96	50	285	85
Percent	100	38	25	12	74	26	38	26	14	77	23

Note: Numbers in parenthesis are number of classifications based on random assignments of subjects to groups in proportion to the total number of subjects in each group.

Table 20. Analysis 2: Classification of criterion sample using the separate dispersion

Training Groups	No. in Sample	Type of Classification									
		Conditional rule					Bayesian rule				
		correct	close	fairly close	Total	Miss	correct	close	fairly close	Total	Miss
1,2 Control sys., Industrial	55	16 (8.2)	12	21	49	6	12 (8.2)	16	16	44	11
3 Natural resources	34	13 (3.1)	6	7	26	8	6 (3.1)	12	6	24	10
4 Engineering	35	13 (3.3)	8	2	23	12	16 (3.3)	5	4	25	10
5 Plans and design	35	1 (3.3)	17	6	24	11	1 (3.3)	16	8	25	10
6 Laboratory	54	28 (7.9)	9	8	45	9	32 (7.9)	6	8	46	8
7 Health	46	30 (5.7)	10	1	41	5	23 (5.7)	17	1	41	5
8 Commercial	56	12 (8.5)	24	2	38	18	21 (8.5)	17	2	40	16
9 Electrical-electronics	55	23 (8.2)	13	4	40	15	27 (8.2)	11	5	43	12
Total	370	136	99	51	286	84	138	100	50	288	82
Percent	100	37	27	14	77	23	37	27	14	78	22

Note: Numbers in parenthesis are number of classifications based on random assignments of subjects to groups in proportion to the total number of subjects in each group.

Table 21. Analysis 2: Classification of validation sample using the common dispersion

Training Groups	No. in Sample	Conditional rule: correct	Conditional rule: close	Conditional rule: fairly close	Conditional rule: Total	Conditional rule: Miss	Bayesian rule: correct	Bayesian rule: close	Bayesian rule: fairly close	Bayesian rule: Total	Bayesian rule: Miss
1,2 Control sys., Industrial	14	4 (2.2	5	1	10	4	8 (2.2)	0	2	10	4
3 Natural resources	8	1 (0.7)	1	1	3	5	0 (0.7)	1	2	3	5
4 Engineering	8	4 (0.7)	1	2	7	1	3 (0.7)	3	0	6	2
5 Plans and design	9	0 (0.9)	4	4	8	1	0 (0.9)	4	2	6	3
6 Laboratory	13	4 (1.9)	2	2	8	5	4 (1.9)	2	2	8	5
7 Health	11	7 (1.3)	2	1	10	1	6 (1.3)	3	1	10	1
8 Commercial	14	3 (2.2)	4	2	9	5	4 (2.2)	5	2	11	3
9 Electrical-electronics	14	7 (2.2)	1	2	10	4	7 (2.2)	4	0	11	3
Total	91	30	20	15	65	26	32	22	11	65	26
Percent	100	33	22	16	71	29	35	24	12	71	29

Note: Numbers in parenthesis are number of classifications based on random assignment of subjects to groups in proportion to the total number of subjects in each group.

Table 22. Analysis 2: Classification of validation sample using the separate dispersion

Training Groups	No. in Sample	Type of Classification: Conditional rule					Bayesian rule				
		correct	close	fairly close	Total	Miss	correct	close	fairly close	Total	Miss
1,2 Control sys., Industrial	14	6 (2.2)	2	4	12	2	5 (2.2)	4	2	11	3
3 Natural resources	8	1 (0.7)	1	1	3	5	0 (0.7)	2	1	3	5
4 Engineering	8	5 (0.7)	1	1	7	1	4 (0.7)	2	1	7	1
5 Plans and design	9	0 (0.9)	3	1	4	5	0 (0.9)	1	3	4	5
6 Laboratory	13	4 (1.9)	4	1	9	4	4 (1.9)	4	1	9	4
7 Health	11	8 (1.3)	1	1	10	1	5 (1.3)	4	1	10	1
8 Commercial	14	2 (2.2)	6	2	10	4	2 (2.2)	6	2	10	4
9 Electrical-electronics	14	5 (2.2)	4	2	11	3	7 (2.2)	2	3	12	2
Total	91	31	22	13	66	25	27	25	14	66	25
Percent	100	34	22	16	71	29	30	24	12	71	29

Note: Numbers in parenthesis are number of classifications based on random assignment of subjects to groups in proportion to the total number of subjects in each group.

procedure of combining groups had some advantage in improving the efficiency of classification, although it could not be overdone at the expense of reducing the number of groups to be classified.

Analysis 3 (9 training groups, 10 predictors). From the results of the first analysis it was found that out of 19 predictor variables, ten of them had discriminant weights of at least .30. A third discriminant analysis was done to verify how well these highly weighted variables would perform in discriminating 9 training groups of NAIT graduates. Ten predictors used in this analysis were the DAT VR, DAT NA, DAT AR, DAT MR, DAT SR, DAT LU II, Lorge-Thorndike (verbal), Math. 9 (Alberta), SCAT 3 (quantitative), and the Grade X average.

Tables 23, 24, 25, and 26 show the results of analysis 3. The first three roots in this analysis were found significant at the .001-level. The observed F-ratio of 5.09 was significant at the .001-level. Again, three discriminant functions were used to determine the dimensionality of the reduced space. The positions of high discriminant weights of these predictors were exactly the same as those found in analysis 1 (refer to Table 7). These results indicated the possibility of using a lesser number of predictors than the 19 variables used in analysis 1.

The results of the classification of the subjects in the criterion and validation samples are shown in Table 27 and 28. Thirty-four percent to 35% of the subject in the criterion sample were correctly classified under both classification rules. The overall predicted classification, close and fairly close classifications included, was 68% to 70% correct.

Table 23. Analysis 3: Roots, Chi square test of dimensionality of discriminant space, and multivariate F-test

Root	λ	DF	Chi square	p	Percent
1	.7948	80	372.26	< .001	61.02
2	.2499	63	166.09	< .001	19.18
3	.1191	48	87.47	< .001	9.14
4	.0634	35	47.82	.073	4.87
5	.0433	24	26.15	.346	3.33
6	.0243	15	11.20	.739	1.86
7	.0049	8	2.73	.950	0.37
8	.0029	3	1.02	.797	0.22
				Total	100.00

Lambda = .348
DF_1 = 80; DF_2 = 2247
F = 5.09
$p < .001$

Table 24. Analysis 3: Centours of group centroids in discriminant space

Group No.	Training Group	Discriminant function I	II	III
1	Control system	-2.54	4.12	2.58
2	Industrial	-2.55	3.80	2.41
3	Natural resources	-2.57	4.42	1.96
4	Engineering	-2.97	4.43	2.04
5	Plans and design	-2.66	4.20	2.59
6	Laboratory	-1.96	4.74	3.39
7	Health	-1.58	4.05	2.29
8	Commercial	-2.22	3.87	2.10
9	Electrical-electronics	-2.72	4.18	2.43

Table 25. Analysis 3: Group centroid vectors in discriminant space

Group centroid	Training group 1	2	3	4	5	6	7	8	9
1	100								
2	94	100							
3	72	64	100						
4	59	51	87	100					
5	99	89	73	70	100				
6	44	24	49	16	37	100			
7	23	22	17	03	14	56	100		
8	73	85	68	33	59	43	57	100	
9	97	90	83	83	99	32	11	62	100

Table 26. Analysis 3: Discriminant weights of three functions

Predictor variables	Discriminant function I	II	III
	(normalized weights)		
1. DAT VR	-.044	.375	-.136
2. " NA	-.543	-.035	-.535
3. " AR	.079	-.346	.219
4. " MR	-.624	.211	.150
5. " SR	.091	-.071	.376
6. " LU II (grammar)	.462	-.171	.322
7. Lorge-Thorndike (V)	-.043	.178	-.351
8. Math. 9 (Alberta)	-.079	.294	-.009
9. SCAT 3 (quantitative)	.168	-.330	.457
10. Grade X average	.223	.657	.218

Table 27. Analysis 3: Classification of criterion sample using the common dispersion

Training Groups	No. in Sample	Conditional rule: correct	Conditional rule: close	Conditional rule: fairly close	Conditional rule: Total	Conditional rule: Miss	Bayesian rule: correct	Bayesian rule: close	Bayesian rule: fairly close	Bayesian rule: Total	Bayesian rule: Miss
1 Control system	30	4 (2.4)	3	11	18	12	1 (2.4)	2	6	9	21
2 Industrial	26	5 (1.8)	4	4	13	13	0 (1.8)	6	7	13	13
3 Natural resources	34	10 (3.1)	11	1	22	12	5 (3.1)	11	6	22	12
4 Engineering	35	17 (3.3)	6	5	28	7	14 (3.3)	12	1	27	8
5 Plans and design	35	10 (3.3)	8	2	20	15	0 (3.3)	13	7	20	15
6 Laboratory	54	30 (7.9)	12	0	42	12	34 (7.9)	8	2	44	10
7 Health	46	28 (5.7)	6	2	36	10	26 (5.7)	9	1	36	10
8 Commercial	56	14 (8.5)	14	8	36	20	19 (8.5)	17	4	40	16
9 Electrical-electronics	55	6 (8.2)	18	13	37	18	31 (8.2)	11	7	49	6
Total	371	124	82	46	252	119	130	89	41	260	111
Percent	100	34	22	12	68	32	35	24	11	70	30

Note: Numbers in parenthesis are number of classifications based on random assignment of subjects to groups in proportion to the total number of subjects in each group.

Table 28. Analysis 3: Classification of validation sample using the common dispersion

Training Groups	No. in Sample	Conditional rule: correct	Conditional rule: close	Conditional rule: fairly close	Conditional rule: Total	Conditional rule: Miss	Bayesian rule: correct	Bayesian rule: close	Bayesian rule: fairly close	Bayesian rule: Total	Bayesian rule: Miss
1 Control system	7	0 (.55)	1	1	2	5	0 (.55)	0	2	2	5
2 Industrial	6	1 (.40)	2	0	3	3	1 (.40)	0	2	3	3
3 Natural resources	8	0 (.71)	2	1	3	5	0 (.71)	1	2	3	5
4 Engineering	8	5 (.71)	1	0	6	2	4 (.71)	1	1	6	2
5 Plans and design	9	4 (.90)	1	3	8	1	0 (.90)	6	2	8	1
6 Laboratory	13	4 (1.9)	3	1	8	5	5 (1.9)	3	1	9	4
7 Health	11	6 (1.3)	3	0	9	2	7 (1.3)	2	0	9	2
8 Commercial	14	2 (2.2)	4	1	7	7	4 (2.2)	4	1	9	5
9 Electrical electronics	14	1 (2.2)	5	3	9	5	8 (2.2)	4	1	13	1
Total	90	23	22	10	55	35	29	21	12	62	28
Percent	100	26	24	11	61	39	32	23	13	69	31

Note: Numbers in parenthesis are number of classifications based on random assignment of subjects to groups in proportion to the total number of subjects in each group.

Predicted classifications of the subjects in the validation sample were 26% to 32% correct. Groups 1 and 3 were not correctly classified but these results were also similar to analysis 1. The overall predicted classifications were 59% to 63% correct. As a whole the prediction in analysis 3 were slightly less accurate than those in analysis 1. Generally there was some merit in considering an analysis involving the ten highly weighted predictors in this study instead of the original nineteen.

Analysis 4 (8 training groups, 10 predictors). It was shown in the second analysis that there was a marked increase in correct classifications when the groups were reduced by combining groups 1 and 2, thus using eight groups instead of the original nine. A fourth analysis was done involving the 10 highly weighted predictors and 8 training groups. This would further verify the merits of the two approaches used.

From the results of this analysis, as shown in Tables 29, 30, 31, and 32, the first three roots were found significant at the .004-level. For testing the difference in group centroids the F-ratio approximation of 5.34 was found to be significant at the .001-level. Three discriminant functions were used to determine the dimensionality of the reduced space. Again the high weights of the ten predictors obtained in the three functions were similar to those in analysis 2.

The results of the classification of subjects in the criterion and validation samples in this analysis are shown in Tables 33 and 34. Both classification rules accounted for 32% to 34% of all the criterion sample correctly predicted. These results, however, were lower than those

Table 29. Analysis 4: Roots, Chi square test of dimensionality of discriminant space, and multivariate F-test

Root	λ	DF	Chi square	p	Percent
1	.8033	70	342.66	< .001	66.39
2	.2097	54	134.52	< .001	17.33
3	.1001	40	67.32	.004	8.27
4	.0453	28	33.65	.213	3.75
5	.0253	18	18.00	.456	2.09
6	.0213	10	9.17	.516	1.76
7	.0049	4	1.72	.787	.41
				Total	100.00

Lambda = .379
DF_1 = 70; DF_2 = 2065
$F = 5.34$; $p < .001$

Table 30. Analysis 4: Discriminant weights of three functions

Predictor variables	Discriminant function I	II (normalized weights)	III
1. DAT VR	.027	.228	-.136
2. " NA	-.509	.013	-.473
3. " AR	.081	-.273	.105
4. " MR	-.657	.270	.093
5. " SR	.111	-.155	.443
6. " LU II (grammar)	.436	-.140	.123
7. Lorge-Thorndike (V)	-.067	.204	-.327
8. Math. 9 (Alberta)	-.065	.371	-.042
9. SCAT 3 (quantitative)	.158	-.404	.473
10. Grade X average	.255	.643	.439

Table 31. Analysis 4: Group centroid vectors in discriminant space

Group No.	Training Group	Discriminant function I	II	III
12	Control sys - industrial	-2.39	4.35	2.08
3	Natural resources	-2.36	4.81	1.67
4	Engineering	-2.76	4.81	1.85
5	Plans and design	-2.44	4.53	2.34
6	Laboratory	-1.71	5.09	2.11
7	Health	-1.36	4.36	2.03
8	Commercial	-2.02	4.25	1.83
9	Electrical-electronics	-2.51	4.53	2.18

Table 32. Analysis 4: Centours of group centroids in discriminant space

Group Centroid	Training group 12	3	4	5	6	7	8	9
12	100							
3	79	100						
4	74	85	100					
5	96	67	73	100				
6	33	45	16	34	100			
7	20	15	03	14	58	100		
8	85	70	37	63	47	56	100	
9	98	79	86	99	30	12	66	100

Table 33. Analysis 4: Classification of criterion sample using the common dispersion

Training Groups	No in Sample	Conditional rule: correct	Conditional rule: close	Conditional rule: fairly close	Conditional rule: Total	Conditional rule: Miss	Bayesian rule: correct	Bayesian rule: close	Bayesian rule: fairly close	Bayesian rule: Total	Bayesian rule: Miss
1,2 Control sys., Industrial	55	6 (8.2)	18	18	42	13	11 (8.2)	25	10	46	9
3 Natural resources	34	14 (3.1)	7	0	21	13	5 (3.1)	11	5	21	13
4 Engineering	35	16 (3.3)	7	2	25	10	12 (3.3)	9	1	22	13
5 Plans and design	35	13 (3.3)	4	4	21	14	1 (3.3)	8	6	15	20
6 Laboratory	54	28 (7.9)	14	3	45	9	33 (7.9)	9	3	45	9
7 Health	46	24 (5.7)	11	1	36	10	28 (5.7)	7	1	36	10
8 Commerical	56	13 (8.5)	15	8	36	20	15 (8.5)	15	9	39	17
9 Electrical-electronics	55	5 (8.2)	30	6	41	14	22 (8.2)	15	8	45	10
Total	370	119	106	42	267	103	127	99	43	269	101
Percent	100	32	29	11	72	28	34	27	12	73	27

Note: Numbers in parenthesis are number of classifications based on random assignment of subjects to groups in proportion to the total number of subjects in each group.

Table 34. Analysis 4: Classification of validation sample using the common dispersion

Training Groups	No. in Sample	Conditional rule: correct	Conditional rule: close	Conditional rule: fairly close	Conditional rule: Total	Conditional rule: Miss	Bayesian rule: correct	Bayesian rule: close	Bayesian rule: fairly close	Bayesian rule: Total	Bayesian rule: Miss
1,2 Control sys., Industrial	14	4 (2.2)	1	5	10	4	5 (2.2)	2	5	12	2
3 Natural resources	8	1 (0.7)	2	0	3	5	0 (0.7)	1	1	2	6
4 Engineering	8	5 (0.7)	0	1	6	2	0 (0.7)	5	1	6	2
5 Plans and design	9	5 (0.9)	1	2	8	1	1 (0.9)	3	2	6	3
6 Laboratory	13	4 (1.9)	4	0	8	5	5 (1.9)	3	0	8	5
7 Health	11	8 (1.3)	1	0	9	2	5 (1.3)	4	0	9	2
8 Commercial	14	3 (2.2)	5	1	9	5	4 (2.2)	4	3	11	3
9 Electrical-electronics	14	1 (2.2)	8	2	11	3	7 (2.2)	2	4	13	1
Total	91	31	22	11	64	27	27	24	16	67	24
Percent	100	34	24	12	70	30	30	26	18	74	26

Note: Numbers in parenthesis are number of classifications based on random assignment of subjects to groups in proportion to the total number of subjects in each group.

obtained in analysis 2. The overall predicted classifications were 72% to 73% correct.

Predicted classifications of subjects in the validation sample were 30% to 34% correct. These results were also lower than analysis 2. The overall predicted classifications were 70% to 74% correct. Analysis 4, involving 8 training groups and 10 predictors produced results which were slightly lower than those of analysis 2, which involved 8 training groups and 19 predictors. If a situation arose where the data of the 19 predictors were not all available, the vocational counselor could use analysis 4 as an alternate predicting scheme for use in counseling.

Analysis 5 (8 training groups, 12 predictors). At the Northern Alberta Institute of Technology, the test scores of students in the Differential Aptitude Tests, the Lorge-Thorndike Intelligence Tests, and the high school averages in Grades X and XI were readily available to the vocational counselor. As mentioned earlier, the vocational guidance and counseling services of NAIT had administered these tests to new students who were applying for admission. Noting that the results of discriminant analysis 4 indicated a little lower results compared with analysis 2, and that most of the sub-tests of these tests stated above were those used in analysis 4, a fifth analysis was carried out involving 8 training groups and 12 predictors to verify the feasibility of an alternate predicting scheme. The 12 predictors used were the eight sub-scores of the DAT, the two sub-scores of the Lorge-Thorndike, and the high school averages in Grades X and XI. These variables included 8 predictors which were used in analysis 4. The same training

groups which were used in analysis 2 were used in this analysis.

Results of analysis 5 are shown in Tables 35, 36, 37, and 38. Again the first three roots were found significant at the .001-level. Three discriminant functions were used to determine the dimensionality of the reduced space. Testing the difference in group centroids, the observed F-ratio of 4.91 was found to be significant at the .001-level. Of the 12 predictor variables used in analysis 5, only the Lorge-Thorndike sub-tests had low discriminant weights in all three functions.

Results of the classification of subjects in the criterion and validation samples are shown in Tables 39 and 40. For subjects in the criterion sample both classification rules accounted for 35% to 37% correct prediction. These results were 3% higher than the results in analysis 4, but 2% lower than those in analysis 2. The overall result of predicted classifications was 74% correct. This was a little higher than the results in analysis 4 but a little lower than those in analysis 2.

Classifications of subjects in the validation sample resulted in 34% to 35% correct prediction. These results were 2% higher than those in analysis 4 but again about 2% lower than those in analysis 2. The overall result of predicted classifications was 64% to 67% correct. This was lower than the results obtained in analysis 2 and 4.

As a whole the classification results in the fifth analysis which used 12 predictors was comparable with those in analysis 4 which used 10 predictors. The prediction based on the fifth analysis was less accurate than those based on analysis 2 which used 19 predictors. There was some advantage in using more predictors in this study.

Table 35. Analysis 5: Roots, Chi square test of dimensionality of discriminant space, and multivariate F-test

Root	λ	DF	Chi square	p	Percent
1	.8865	84	377.02	< .001	65.79
2	.1833	66	153.60	< .001	13.60
3	.1087	50	94.36	< .001	8.07
4	.0733	36	58.03	.011	5.44
5	.0390	24	33.12	.101	2.89
6	.0337	14	19.66	.141	2.50
7	.0231	6	8.01	.237	1.71
				Total	100.00

Lambda = .343

$DF_1 = 84$; $DF_2 = 2158$

$F = 4.91$; $p = < .001$

Table 36. Analysis 5: Discriminant weights of three functions

Predictor variables	Discriminant function I	II (Normalized weights)	III
1. DAT VR	.009	.226	.161
2. " NA	-.515	.108	.208
3. " AR	.199	-.312	.141
4. " CSA	.196	-.015	.228
5. " MR	-.660	.221	.331
6. " SR	.096	-.109	-.316
7. " LU I (spell)	.133	-.328	.164
8. " LU II (grammar)	.369	.046	-.246
9. Lorge-Thorndike (V)	-.037	.116	.051
10. " " (NV)	-.167	.105	-.103
11. Grade X average	.096	.796	-.543
12. Grade XI average	.143	.102	.503

Table 37. Analysis 5: Group centroid vectors in discriminant space

Group No.	Training Group	Discriminant function I	II	III
12	Control sys - industrial	-1.06	2.31	3.81
3	Natural resources	-1.01	2.48	3.81
4	Engineering	-1.48	2.62	3.77
5	Plans and design	-1.17	2.37	3.61
6	Laboratory	-0.28	2.84	3.77
7	Health	0.05	2.35	3.52
8	Commercial	-0.66	2.15	3.76
9	Electrical-electronics	-1.40	2.56	3.47

Table 38. Analysis 5: Centours of group centroids in discriminant space

Group Centroid	Training groups 12	3	4	5	6	7	8	9
12	100							
3	98	100						
4	79	83	100					
5	95	93	86	100				
6	35	50	16	28	100			
7	19	22	03	16	63	100		
8	88	83	36	74	49	53	100	
9	72	74	90	93	16	06	38	100

Table 39. Analysis 5: Classification of criterion samples using the common dispersion

Training Groups	No. in Sample	Conditional rule: correct	Conditional rule: close	Conditional rule: fairly close	Conditional rule: Total	Conditional rule: Miss	Bayesian rule: correct	Bayesian rule: close	Bayesian rule: fairly close	Bayesian rule: Total	Bayesian rule: Miss
1,2 Control sys., Industrial	55	13 (8.2)	11	12	36	19	20 (8.2)	12	12	44	11
3 Natural resources	34	4 (3.1)	7	14	25	9	0 (3.1)	6	6	12	22
4 Engineering	35	15 (3.3)	7	4	26	9	8 (3.3)	13	2	23	12
5 Plans and design	35	6 (3.3)	10	7	23	12	0 (3.3)	10	10	20	15
6 Laboratory	54	25 (7.9)	14	7	46	8	29 (7.9)	12	5	46	8
7 Health	46	29 (5.7)	12	0	41	5	28 (5.7)	13	0	41	5
8 Commercial	56	14 (8.5)	19	6	39	17	19 (8.5)	20	5	44	12
9 Electrical-electronics	55	25 (8.2)	7	6	38	17	33 (8.2)	5	6	44	11
Total	370	131	87	56	274	96	137	91	46	274	96
Percent	100	35	24	15	74	26	37	25	12	74	26

Note: Numbers in parenthesis are number of classifications based on random assignment of subjects to groups in proportion to the total number of subjects in each group.

Table 40. Analysis 5: Classification of validation sample using the common dispersion

Training Group	No. in Sample	Conditional rule: correct	Conditional rule: close	Conditional rule: fairly close	Conditional rule: Total	Conditional rule: Miss	Bayesian rule: correct	Bayesian rule: close	Bayesian rule: fairly close	Bayesian rule: Total	Bayesian rule: Miss
1,2 Control sys., Industrial	14	4 (2.2)	2	1	7	7	5 (2.2)	4	1	10	4
3 Natural resources	8	1 (0.7)	1	0	2	6	0 (0.7)	0	1	1	7
4 Engineering	8	5 (0.7)	2	0	7	1	3 (0.7)	4	0	7	1
5 Plans and design	9	2 (0.9)	0	5	7	2	0 (0.9)	0	7	7	2
6 Laboratory	13	6 (1.9)	2	0	8	5	6 (1.9)	1	1	8	5
7 Health	11	7 (1.3)	2	0	9	2	5 (1.3)	4	0	9	2
8 Commercial	14	3 (2.2)	1	3	7	7	3 (2.2)	2	3	8	6
9 Electrical-electronics	14	4 (2.2)	5	2	11	3	9 (2.2)	2	0	11	3
Total	91	32	15	11	58	33	31	17	13	61	30
Percent	100	35	17	12	64	36	34	19	14	67	33

Note: Numbers in parenthesis are number of classifications based on random assignment of subjects to groups in proportion to the total number of subjects in each group.

Cross-classification on non-graduates

The prediction scheme developed in this study was applied on a sample of non-graduates to verify how well it would classify a different population category. As shown in Table 3, there were 403 non-graduates who were not included in the previous analyses. Centours, Bayesian probabilities, and classifications were computed on all individuals in this sample. The prediction weights obtained in analysis 2 based on 8 training groups and 19 predictors were used in the classification solution.

Certainly the expectation was that non-graduates would be misclassified. However, this might not be so for all of this type of student. The causes of students withdrawing from school or transferring to another course may not be similar for all students. A student may withdraw from school not because he lacks aptitude for the training course but, perhaps because the course lacks challenge. Possibly he could have graduated had he stayed and finished the course.

The results of the cross-classification are shown in Table 41. Twenty-two percent to 23% of the subjects in the non-graduate sample were correctly classified while 36% to 39% were missed entirely. These results were opposite in trend to the results obtained in analysis 2 (refer to Tables 19 and 21). In analysis 2 the predicted classifications of graduates were 33% to 38% correct while 22% to 26% were misclassified.

The cross-classification result indicated that the prediction scheme developed in this study could classify about half of the non-graduate population who were likely to lack aptitude and abilities

Table 41. Cross-classification of non-graduate sample using the discriminant weights and common dispersion in analysis 2 (graduate criterion sample)

	Training Groups	NO.	Type of Classification: Conditional rule correct	close	fairly close	Total	Miss	Bayesian rule correct	close	fairly close	Total	Miss
1,2	Control sys., industrial	54	6	10	17	33	21	10	20	9	39	15
3	Natural resources	12	3	2	0	5	7	1	1	3	5	7
4	Engineering	18	1	5	3	9	9	1	5	2	8	10
5	Plans and design	45	7	12	13	32	13	0	9	17	26	19
6	Laboratory	44	11	9	2	22	22	13	8	1	22	22
7	Health	19	10	6	0	16	3	7	9	0	16	3
8	Commercial	100	25	42	6	73	27	27	42	9	78	22
9	Electrical-electronics	111	24	17	16	47	54	32	24	9	65	46
	Total	403	87	103	57	247	156	91	118	50	259	144
	Percent	100	22	26	14	61	39	23	29	12	64	36

for technical training. However, to identify non-graduates as such was another problem which was dealt with in a later analysis. From the verifications carried out, the results indicated that the prediction scheme developed here could be a useful tool for vocational counseling at NAIT.

Psychological interpretations

Eight instead of nine training groups in discriminant analysis 2, showed a fairly good number of correct classifications. The results of analysis 2 were further used in examining how the scores of the NAIT graduate criterion sample were distributed in the three-dimensional discriminant space. The location of the 8 group centroids in this reduced space as shown in Figure 3 was plotted as projections along each discriminant function from low to high and is shown in Figure 4. The purpose was to derive a psychological meaning which might be attributed to each of the functions. Along discriminant function I, the group centroids were found to arrange from low to high as follows:

Low - Health
- Laboratory
- Commerical
- Industrial-control systems
- Natural resources
- Plans and design
- Electrical-electronics
High - Engineering

As examination of the arrangement of the groups indicated that those at the high end were those requiring mechanical skills and abilities. Discriminant function I may be labeled as mechanical attribute. Health, laboratory, and commercial training groups probably did not require as much of the mechanical attribute as did the electrical and

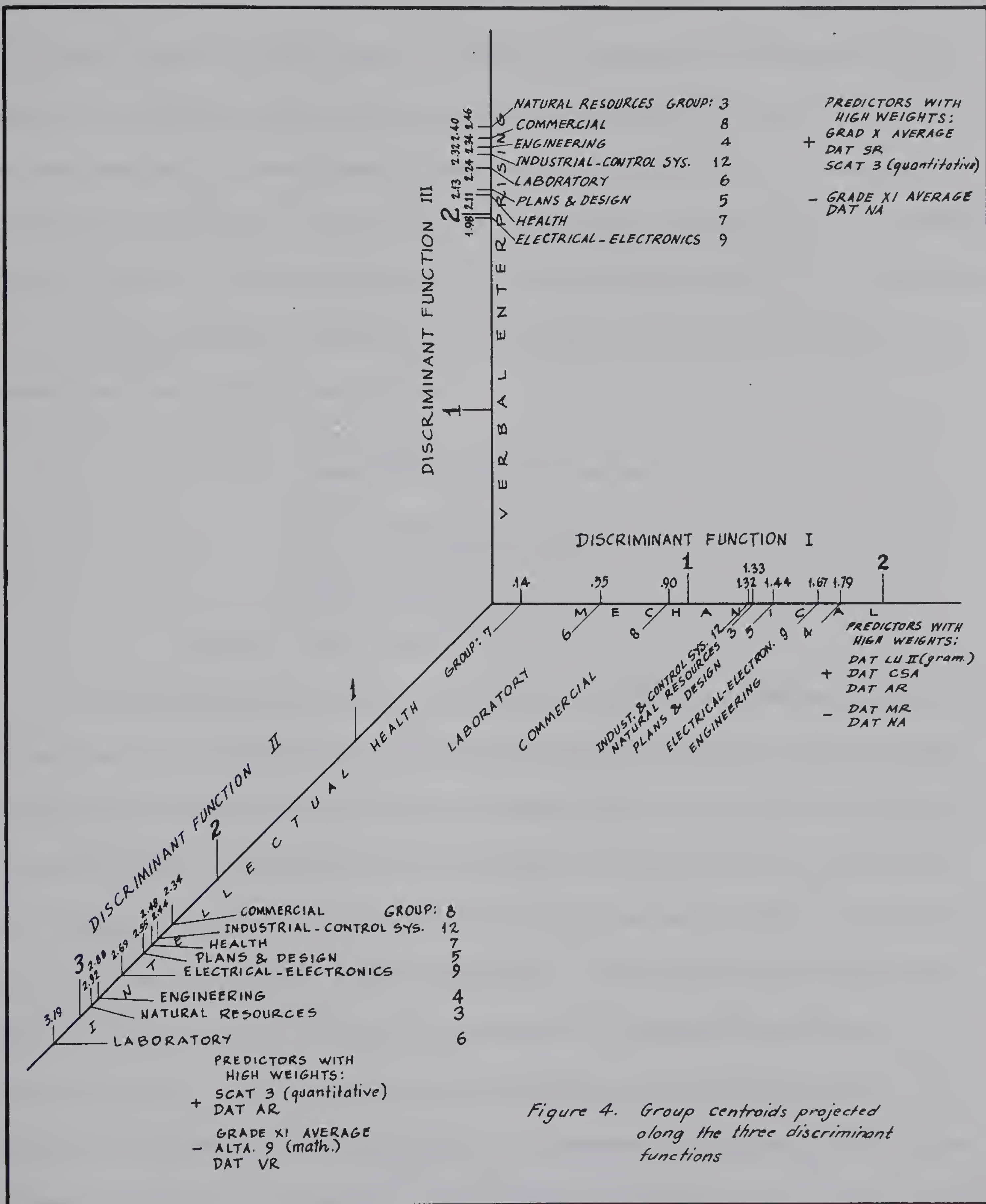

Figure 4. Group centroids projected along the three discriminant functions

engineering groups. In function I, the DAT LU II(grammar), DAT CSA, and DAT AR were the variables weighted highly positive while DAT MR and DAT NA were the variables weighted highly negative. This mechanical attribute might be interpreted in terms of measures of mechanical and numerical abilities (represented by the DA MR and DAT NA variables) on one extreme end of the graduate population and measures of language, clerical, and abstract reasoning (represented by the DAT LU II - grammar, DAT CSA, and DAT AR) variables on the other extreme end of the population.

In discriminant function II the group centroids were found to range from low to high as follows:

Low - Commercial
- Industrial-control systems
- Health
- Plans and design
- Electrical-electronics
- Engineering
- Natural resources

High - Laboratory

This psychological attribute could probably be labeled as intellectual. Laboratory and engineering groups probably require higher intellectual abilities than those in commercial and industrial-control systems groups. In function II, the SCAT 3 (quantitative) and DAT AR were weighted highly positive while Grade XI average, Math. 9 (Alberta), and DAT VR were weighted highly negative. The intellectual attribute might be interpreted in terms of measures of general intelligence (represented by SCAT 3 and DAT AR) variables on one extreme end of the population and measures of scholastic achievement (represented by Grade XI average, Math. 9 (Alberta), and DAT VR) variables on the other extreme end of the population.

In discriminant function III, the group centroids were found to range from low to high as follows:

Low - Electrical-electronics
- Health
- Plans and design
- Laboratory
- Industrial-control systems
- Engineering
- Commercial

High - Natural resources

This psychological attribute could probably be labeled as verbal-enterprising. The electrical-electronics and plans and design groups probably require less of the verbal-enterprising attribute than those in the commercial and natural resources groups. In function III, Grade X average, DAT SR, SCAT 3 (quantitative) were weighted highly positive while Grade XI and DAT NA were weighted highly negative. The verbal-enterprising attribute might be interpreted in terms of measures of scholastic achievement, general intelligence, and imagination (represented by Grade X average, DAT SR, SCAT 3 - quantitative) variables versus measures of numerical ability and scholastic achievement (represented by Grade XI average and DAT NA) variables.

Although the predictor variables used were limited to areas of aptitudes, intelligence, and scholastic achievement, it was interesting to note how discriminant analysis in this study could distinguish dimensions of group attributes with some degree of precision. The differentation of attributes shown in this study suggested that the use of more predictors which measures various psychological characteristics would probably achieve more distinct identification of groups than the measures used here.

Analysis 6 (2 outcome categories, 19 predictors). This analysis investigated whether the 19 predictors used in this study could differentiate between those who would be expected to graduate and those who would not. The criterion samples were 323 graduates and 282 non-graduates.

The results of the analysis of variance for testing the significance of group means and homogeneity of variance of the 19 predictors are shown in Table 42. Group means in ten predictors were not significant at the .05-level while nine were significant at the .002-level. Hypothesis (1) of no difference in group means was not entirely rejected. This preliminary analysis showed that 10 of the variables would not differentiate effectively between graduates and non-graduates.

The chi square values of 12 predictors were significant at $.06 > p > .87$ except for 7 predictors which were significant at $.04 > p > .001$. The hypothesis of homogeneity of variance was not entirely tenable. These results indicated the possibility of large overlap between groups and thereby increased misclassification of persons.

The intercorrelations of the 19 variables were also computed and the results are shown in Table 43. Out of 171 pairs of correlations 28 pairs obtained correlations over .50. These pairs were:

1) DAT SR and DAT AR, DAR MR, Lorge-Thorndike (NV);
2) Lorge-Thorndike (V) and DAT VR, LU II (grammar), Read.-lit. 9 (Alberta), SCAT 3 (V);
3) Lorge-Thorndike (NV) and DAT AR;
4) Read.-lit. 9 (Alberta) and DAT VR, LU II (grammar), Science 9 (Alberta), SCAT 3 (V);
5) Lang. 9 (Alberta) and LU II (grammar), SCAT 3 (V), Grade X average,

Table 42. Analysis 6: Means, univariate F-tests and Chi square tests of homogeneity of variance

Predictors	Group means[b] 1	2	F^a	p	Chi Sq.	p
1 DAT VR	3.71	3.62	2.72	.099	1.75	.186
2 " NA	3.48	3.39	6.99	.008	12.70	<.001
3 " AR	3.96	3.94	.08	.783	10.37	.001
4 " CSA	5.43	5.29	3.29	.070	1.14	.285
5 " MR	5.44	5.35	1.84	.175	3.49	.062
6 " SR	3.90	3.87	.11	.738	.05	.816
7 " LU I (spell.)	8.40	8.43	.19	.661	4.80	.028
8 " LU II (gram.)	3.78	3.73	.60	.440	.96	.327
9 Lorge-Thorn. (V)	6.15	6.01	3.77	.053	.03	.868
10 " " (NV)	5.08	4.99	1.55	.213	7.19	.007
11 Read.-lit.9 (Alta.)	6.12	5.81	9.77	.002	.72	.396
12 Lang. 9 (Alta.)	6.17	5.80	12.71	<.001	.13	.716
13 Soc. stu. 9 (Alta.)	6.56	6.21	13.54	<.001	4.26	.039
14 Math. 9 (Alta.)	6.58	6.18	14.86	<.001	4.44	.035
15 Science 9 (Alta.)	6.73	6.33	17.40	<.001	8.35	.004
16 SCAT 3 (V)	4.18	4.06	2.35	.125	.71	.396
17 SCAT 3 (quant.)	3.78	3.62	6.67	.010	.04	.837
18 Grade X average	3.53	3.28	33.78	<.001	.13	.712
19 Grade XI average	3.34	3.06	34.91	<.001	.63	.427

[a] $DF_1 = 1$; $DF_2 = 603$

[b] 1 - Graduate sample

2 - Non-graduate sample

Table 43. Analysis 6: Intercorrelation and criterion means

	Predictor variables																			Criterion
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	Means
1	1.0																			3.67
2	.42	1.0																		3.44
3	.47	.46	1.0																	3.95
4	.04	.15	.14	1.0																5.36
5	.31	.25	.37	-.15	1.0															5.40
6	.37	.32	.54	.09	.58	1.0														3.88
7	.23	.16	.04	.28	-.16	-.10	1.0													8.42
8	.47	.29	.23	.19	.07	.12	.47	1.0												3.75
9	.67	.38	.37	.03	.34	.35	.29	.51	1.0											6.09
10	.42	.46	.62	.19	.37	.56	.13	.22	.43	1.0										5.04
11	.59	.29	.26	.14	.05	.11	.40	.54	.64	.21	1.0									5.98
12	.45	.31	.21	.22	-.06	.02	.49	.55	.46	.21	.67	1.0								5.99
13	.43	.33	.18	.06	.12	.13	.23	.32	.52	.18	.61	.53	1.0							6.40
14	.37	.51	.32	.15	.15	.21	.19	.32	.38	.33	.42	.49	.52	1.0						6.39
15	.43	.43	.28	.05	.28	.25	.12	.30	.48	.25	.51	.47	.69	.63	1.0					6.55
16	.56	.22	.19	.002	.19	.14	.30	.48	.65	.21	.72	.50	.53	.32	.44	1.0				4.13
17	.35	.44	.33	.20	.11	.19	.23	.26	.36	.37	.43	.41	.38	.57	.42	.47	1.0			3.71
18	.27	.33	.15	.25	-.12	-.004	.33	.35	.28	.13	.43	.56	.51	.51	.50	.25	.33	1.0		3.42
19	.27	.26	.10	.24	-.12	.01	.32	.34	.29	.11	.40	.50	.45	.39	.38	.25	.27	.71	1.0	3.21

Legend:

Predictior
1. - DAT VR
2. - " NA
3. - " AR
4. - " CSA
5. - " MR
6. - " SR
7. - " LU I (spell.)
8. - " LU II (grammar)
9. - Lorge-Thorndike (V)
10. - Lorge-Thorndike (NV)
11. - Read.-lit. 9 (Alberta)
12. - Lang. 9 (Alberta)
13. - Soc. stud. 9 (Alberta)
14. - Math. 9 (Alberta)
15. - Science 9 (Alberta)
16. - SCAT 3 (V)
17. - SCAT 3 (NV)
18. - Grade X average
19. - Grade XI average

Grade X average;

7) Math. 9 (Alberta) and DAT NA, Soc. stud. 9 (Alberta), Science 9 (Alberta), SCAT 3 (NV), Grade X average;

8) SCAT 3 (V) and DAT VR;

9) Grade X average and Science 9 (Alberta), Grade XI average.

As a whole the intercorrelations were low. Positive correlations varied from .002 to .71 while negative correlations ranged from -.004 to -.16. The high correlations of these pairs were likely due to the similarity of the characteristics being measured by these variables. This also indicated that some variables were duplicated. The intercorrelations showed that the predictors might differentiate two outcome categories of NAIT students.

The single root obtained in the discriminant analysis as shown in Table 44 was 0.11 which was much lower than the significant roots obtained in the previous analyses involving 9 and 8 training groups. The value indicated that the variance along the discriminant axis was small and might not provide sufficient differentiation between two groups.

However, in testing the difference in group centroids the observed F-ratio of 3.39 was significant at the .001-level. The generalized multivariate null hypothesis (2), that the characteristics of the two outcome categories as measured by the 19 predictors were the same, was rejected. There was difference between the two group centroids and this indicated that differentiation between the two categories might be possible.

The value of the group centroids of the graduate and non-graduate samples along the discriminant axis were 3.36 and 2.96 respectively. The

Table 44. Analysis 6: Root, Chi square test of dimensionality of discriminant space, multivariate F-test, and discriminant weights in one function

Root	λ	DF	Chi sq	p
1	.1110	19	61.95	< .001

Wilks lambda	DF_1	DF_2	F	p
.900	19	585	3.39	< .001

Predictors	Discrim. function I (normalized weights)	Standard deviation	
		Graduates	Non-grads.
1. DAT VR	-.055	.649	.701
2. " NA	.144	.378	.465
3. " AR	-.345	.504	.606
4. " CSA	.112	.893	.950
5. " MR	.358	.750	.835
6. " SR	-.131	1.053	1.067
7. " LU I (spell.)	-.270	1.055	.929
8. " LU II (grammar)	-.202	.771	.728
9. Lorge-Thorndike (V)	-.001	.887	.896
10. " " (NV)	.118	.789	.921
11. Read.-lit. 9 (Alberta)	.175	1.144	1.201
12. Lang. 9 (Alberta)	.090	1.277	1.304
13. Soc. stud. 9 (Alberta)	-.047	1.109	1.250
14. Math. 9 (Alberta)	.018	1.193	1.347
15. Science 9 (Alberta)	.019	1.087	1.285
16. SCAT 3 (V)	-.090	.998	.950
17. SCAT 3 (quantitative)	.038	.729	.738
18. Grade X average	.467	.526	.537
19. Grade XI average	.549	.582	.556

variables which obtained high weights were Grade X average, Grade XI average, DAT MR, DAT AR, DAT LU I (spell.) and DAT LU I (grammar). These group centroid values might be interpreted as technical training success continuum along the discriminant axis as measured by the predictors and represented by the variables above.

The centour score of the graduate group with respect to the non-graduate group shown in Table 45 was 51. This meant that the graduate group overlapped with the non-graduate group by 51% and that about half of both categories would likely be misclassified.

Discriminant scores, centours, Bayesian probabilities, and classifications using the common dispersion were computed for 259 subjects in the validation sample. The results of the classifications are summarized in Table 46. Of the 138 graduates in the validation sample both classification rules correctly predicted 50% to 59% as graduates and 41% to 59% as non-graduates. Of the 212 non-graduates in the validation sample, both classification rules correctly predicted 55% to 66% as non-graduates and 34% to 45% as graduates. The overall predicted classifications were 57% to 58% correct and 42% to 43% misclassified. The low percentage of correct classifications showed that it was not possible to differentiate between graduates and non-graduates with sufficient efficiency in this analysis.

Before the computations of discriminant scores, centours, Bayesian probabilities, and classifications were carried out, the preliminary univariate F-tests and homogeneity of variance tests, the low value of the root and the large overlap obtained already preclude the difficulties of obtaining adequate differentiation between groups. Perhaps the reason for

Table 45. Analysis 6: Group centroids and centours in discriminant space

Group Category	Discrminant function I
	Group centroid
Graduates	3.36
Non-graduates	2.95

Group centroid of	Centour Graduates	Centour Non-graduates
Graduates	100	51
Non-graduates	51	100

Table 46. Analysis 6: Classification of validation sample of graduates and non-graduates using the common dispersion

Group category	No. of Samples	Type of Classification			
		Conditional		Bayesian	
		correct	miss	correct	miss
Graduates	138	69 (50%)	69 (50%)	82 (59%)	56 (41%)
Non-graduates	121	80 (66%)	41 (34%)	66 (55%)	55 (45%)
Total	259	149	110	148	111
Percent	100%	58%	42%	57%	43%

these difficulties was that the population samples used were already preselected by the usual admission procedures of the institute and were, therefore, a restricted group. Another reason was that the causes of non-graduation are not the same for all students. A student may not graduate not because he lacks aptitude and ability for the training but perhaps because the course lacks challenge or because he found a job. He could have graduated had he stayed and finished the course.

Possibly the admission procedures at NAIT were performing a satisfactory job of distinguishing graduates from non-graduates. However, the results of analysis of 8 training groups and 19 predictors in this study would likely be helpful in counseling the student's choice of training course to pursue and supplement the selection procedures of the institute.

Summary

This study investigated the possibility of developing a procedure for predicting training outcomes of graduates at NAIT by using 19 predictor variables. As a preliminary step, a one-way analysis of variance was done to ascertain the usefulness of each variable as a predictor. For each variable, the F-tests of the difference of group means for the 9 training groups of the criterion sample all were found significant at the .013-level. The results showed that each training group was different from the other, as measured by the variables.

Another preliminary verification was made by examining the intercorrelations of the 19 variables. It was found that the intercorrelations were generally low.

The F-ratio approximations, to test the difference of group centroids in the test space were found significant at the .001-level in all discriminant analyses done in this study.

Five discriminant analyses and classification solutions were carried out on the graduate sample. Analysis 1 was based on 9 training groups and 19 predictors. Correct classifications in each training group of the criterion sample accounted for 36% correct prediction. The overall predicted classification was 73% correct. Correct classifications in each training group for the validation sample accounted for 25% correct prediction. The overall predicted classification was 61% correct. The classifications were computed using the total or common dispersion as well as using the separate dispersion in each group and the results in both solutions were found comparatively similar.

Analysis 2 was done, involving 8 training groups and 19 predictors, to examine the possibility of improving the number of correct classifications. The 9 training groups were reduced to 8 groups by combining groups 1 and 2, which were found in the first analysis to overlap to a large extent. Correct classification for the criterion sample increased to 38% correct prediction. The overall predicted classifications increased to 76% correct. For the validation sample the number of correct classifications increased to 33% correct prediction. The overall predicted classifications also increased to 72% correct. The procedure of combining similar groups which overlapped considerably was found to improve correct classification. Again classifications computed using the separate dispersion in each group were found similar to the results computed using the common dispersion of all groups. In subsequent analyses the classification solution was computed using the common dispersion.

Analysis 3 was carried out, involving 9 training groups and 10 predictors which were found highly weighted in the first analysis, to verify the results of classification when the number of predictors were reduced. Correct classifications for the criterion samples accounted for 34% correct prediction. The overall predicted classifications were 69% correct. Correct classifications of the validation samples accounted for 29% correct prediction. The overall predicted classifications were 65% correct. As a whole these results were slightly lower than analysis 1.

Analysis 4 was done involving 8 training groups and 10 highly weighted predictors to verify the results of classification when groups 1 and 2 were combined into one group. Predicted classifications of the criterion samples accounted for 33% correct. The overall predicted classifications increased to 72% correct. In the validation samples the number of correct classifications increased to 32% correct prediction. The overall predicted classifications were 72% correct. The results of this fourth analysis indicated the possibility of using a fewer number of predictors than 19 although prediction would probably be less accurate.

At NAIT, the scores of the DAT, Lorge-Thorndike Intelligence Tests, and the high school averages in Grades X and XI were readily available to the vocational counselor. In the previous analyses, almost all of these variables were found highly weighted. Analysis 5 was carried out involving 8 training groups and the 12 predictors mentioned above. Correct prediction of the criterion sample accounted for 36% correct. The overall predicted classifications were 74% correct.

In the validation sample, the number of correct classifications accounted for 34% correct prediction. The overall predicted classifications were 65% correct. The results of the fifth analysis showed that using 8 groups and 12 variables performed fairly well as compared with analysis 4, but somewhat lower compared with analysis 2.

The prediction weights in analysis 2 were used on subjects in the non-graduate sample to verify how far it would classify a different population category. The results were just the opposite to those obtained in analysis 2. The percentages were similar but in reverse order. The cross-classification results further showed that the prediction scheme developed in this study could be a useful tool for vocational counseling at NAIT. An attempt was made to interpret the meaning of the distribution of scores of the NAIT graduate criterion sample in the three-dimensional discriminant space. The first discriminant function was interpreted as measuring mechanical characteristics of the students, while the second function was measuring intellectual characteristics, and the third function as measuring verbal enterprising characteristics.

As a whole the results showed that the 19 predictor variables could possibly be used in predicting training outcomes of graduates at NAIT with fairly good results. Twelve highly weighted variables in this study could probably perform as well as the 19 predictors but prediction efficiency might be a little less.

A sixth analysis was carried out involving two outcome categories and 19 predictors to investigate whether it was possible to differentiate adequately between graduates and non-graduates. The results showed that this analysis correctly predicted 58% and misclassified 42% of the validation sample. It was concluded that this analysis was not efficient

enough to differentiate between graduates and non-graduates. One reason given was that the samples used were already preselected and were, therefore, a restricted group.

CHAPTER VI

DISCUSSION

Summary

In this study, nineteen predictor variables were used to develop a procedure for predicting training outcomes of graduates of the Northern Alberta Institute of Technology. The problem was to determine whether a combination of 19 predictor variables could discriminate 9 training groups of graduates. Furthermore, the study investigated various modifications which might improve the effectiveness of classification of persons to different training groups.

The discriminant analysis was used to determine the reduced space in which the separation of training groups was maximum. To determine the probability that a person, having a particular set of scores, was a member of a certain training group, the centour classification solution was computed for that person. The effectiveness of classification was verified from a stratified random sample of graduates whose scores were set aside for this purpose at the beginning of this study. It was also verified from the criterion sample which was used in the analysis. The results of the classification were tabulated in three categories: correct, close, and fairly close classifications. Both conditional and Bayesian rules of classification were applied to determine the training group to which a person would most probably belong. In evaluating the effectiveness of classification, comparisons were made in terms of percentages of predicted correct, close, fairly close, and overall classifications, and misclassifications.

Five discriminant analyses and classification solutions were carried out on the graduate sample. In this study, analysis 2 involving 8 training groups and 19 predictor variables, produced the best results of classification.

The effectiveness of analysis 2 was further verified by applying the prediction weights of this analysis to a sample of non-graduates. The results were opposite to those in analysis 2. This further indicated that the prediction scheme developed in this study could be a useful tool for vocational counseling at NAIT.

Several modifications of the analysis were made, such as reducing the number of training groups to minimize large overlaps of groups, limiting the number of predictor variables to those which were weighted high, and carrying out analyses using both limitations. Reducing the number of training groups was found to improve the number of correct classifications but it did not always improve accuracy. Limiting the predictor variables to those which were highly weighted showed lower correct classifications than using all the predictor variables. Analysis was done on 12 predictors which were readily available to vocational counselors at NAIT. The results were also lower than using 19 variables but would probably serve the purpose of vocational counseling although with a lesser degree of effectiveness.

The question whether the 19 predictors could differentiate between graduates and non-graduates was investigated. The results showed that it was not possible to distinguish between these two categories accurately enough for use in counseling.

Interpretations

The nineteen predictor variables did discriminate 8 training groups fairly well. However, it was possible in this study to reduce the number of predictors by using those with high discriminant weights and still obtain fair results. One analysis showed that prediction based on 10 predictor variables were only slightly less accurate than those based on the entire 19 variables. This indicated that some variables were somewhat duplicated and were measuring characteristics similar to the other variables. From the discriminant analysis one might be able to give a parsimonous interpretation of what was being measured by the 19 variables. In this study, the first discriminant function was interpreted as measuring the mechanical domain, while the second function the intellectual domain, and the third function the verbal-enterprising domain. Much improvement could still be made in the selection of predictor variables over those used in this study. Predictor variables which measured specific aptitudes of training fields or training courses and other attributes would probably further increase the discrimination of groups.

The large overlaps of groups 1, 2, 3, 5, and 9 in the first analysis indicated that misclassifications would occur in these groups. This was confirmed by the results of later analyses and classifications. Considering the general aptitude and training requirements in groups 1, 2, 3, 4, and 5 which were largely mechanical, it would appear that these groups were somewhat similar. Could distinction of these groups be further improved? Further investigation on this question was needed using other predictors that could measure certain specific attributes of persons belonging to these training fields and perhaps secure a much clearer

differentiation than this study was able to achieve. As stated earlier, the procedure of combining groups could not be overdone without reducing the usefulness of the prediction scheme for counseling.

When groups 1 and 2 were combined in the second analysis, the overlapping of groups was reduced slightly, but overlaps still occurred in groups 3 and 5. Again the results showed that these two groups were not well classified. The vocational counselor should perhaps be cautious in using the classifications named under these two groups.

The decision on the *a priori* grouping of training fields and the kinds of predictor variables to be used were important factors in this prediction scheme. Each group should be distinct, as measured by the predictor variables in order to secure less overlap among groups. Also, it was important that the sample size in each group, both those used in the analysis and in the validation, should be large enough to insure normality and homogeneity of groups.

In the conditional rule of classification, a person was classified into a certain group when the centour or probability he obtained for that group was the highest among the groups. The vocational counselor should exercise some caution in interpreting such classifications. Cooley and Lohnes (1962) pointed out that this rule assumed that membership in one group occurred as often as that of other groups. When membership in one group occurred more often than the second, this rule tended to classify persons to the second group. The rule also assumed that the dispersions in each group were equal. When the dispersion of one group was greater than that of a second group, this rule tended to classify persons into the first group. However, the classification results

computed using the total dispersion and using the separate dispersion in each group was found to give similar results in this study. This indicated that dispersions of groups in this study were somewhat similar.

The Bayesian classification rule identified the group to which a person most probably belonged. The solution under this rule considered and minimized the possibilities of inequalities of group membership and group dispersion thereby minimizing misclassification. Therefore, it was a better rule to apply than the conditional rule. However, as noted in this study, if the counselor considers the classification obtained by applying both rules he cannot be far from correct in naming the training group to which a student might probably belong.

Centours and probabilities for counseling

Chart 1 is a sample computer print-out of the classification taken in analysis 2 (8 training groups, 19 predictors). The six-digit number (240603) at the top of the left column represents the student ID (24), the specific technology for which the student applied (06), and the training field initially chosen (03). The student's discriminant vector scores in the three functions are shown on the first line. Chi square values of this discriminant score vector, for each training group, are shown on the second line. Posteriori probabilities of each training group, based on the Bayesian rule, are displayed on the third line while the fourth line shows the conditional probabilities obtained from the chi square values.

The highest posteriori probability for student No. 24 is 0.22. Therefore, by the Bayesian classification rule, this student resembled graduates of training field 1 (industrial-control sys.). His highest

Chart 1. Sample Classification print-out of the computer (taken from analysis 2)

	Function	1 I	2 II	3 III	4	5	6	7	8
					Training Groups				
240603	SCORE	-1.5695	-2.7795	-2.5266					
	CHISQ	0.7840	0.2797	0.3837	1.1191	4.3441	7.8460	2.1565	1.7250
	POSTERIOR P	0.2253	0.1792	0.1751	0.1213	0.0373	0.0055	0.1155	0.1407
	CONDITION P	0.8533	0.9638	0.9436	0.7725	0.2266	0.0493	0.5406	0.6314
.....CLASSIFICATION.....		CONDITIONAL:		2	BAYSIAN RULE:	1			
250603	SCORE	-1.9910	-3.7947	-2.4928					
	CHISQ	7.4018	3.8519	2.9366	6.6708	8.2347	17.3441	10.6181	5.6743
	POSTERIOR P	0.0679	0.2477	0.4029	0.0623	0.0440	0.0004	0.0138	0.1611
	CONDITION P	0.0601	0.2779	0.4015	0.0832	0.0414	0.0006	0.0140	0.2386
.....CLASSIFICATION.....		CONDITIONAL:		3	BAYSIAN RULE:	3			
260603	SCORE	-2.0885	-2.6063	-2.3520					
	CHISQ	1.9825	2.2707	0.5150	1.6529	8.7808	12.6242	4.8066	1.3702
	POSTERIOR P	0.1903	0.1019	0.2523	0.1428	0.0062	0.0008	0.0472	0.2585
	CONDITION P	0.5761	0.5182	0.9156	0.6475	0.0324	0.0055	0.1865	0.7125
.....CLASSIFICATION.....		CONDITIONAL:		3	BAYSIAN RULE:	8			
270603	SCORE	-1.7572	-4.1300	-2.6625					
	CHISQ	10.2913	5.4579	5.5812	9.8906	8.5381	18.7787	12.9475	9.2085
	POSTERIOR P	0.0506	0.3505	0.3392	0.0393	0.1193	0.0006	0.0136	0.0869
	CONDITION P	0.0162	0.1412	0.1339	0.0195	0.0361	0.0003	0.0048	0.0266
.....CLASSIFCATION.....		CONDITIONAL:		2	BAYSIAN RULE:	2			

Training group 1 - Industrial-controlsys.
2 - Natural resources
3 - Engineering
4 - Plans and design
5 - Laboratory
6 - Health
7 - Commerical
8 - Electrical-electronics

conditional centour (probability) is 0.96. By the conditional rule this student resembled graduates of training field 2 (natural resources). If the initial choice of this student was training field 3 (engineering) then his group 3 posteriori probability of 0.18 was fairly close to the Bayesian classification of 0.23 (group 1) and his group 3 probability of .94 was close to the conditional classification of 0.96 (group 2).

The vocational counselor may therefore attempt to counsel the student whose initial choice is training field 3 (engineering) by explaining:

1) that he resembles less closely students who satisfactorily completed training field 3 (engineering);

2) that he resembles students who satisfactorily completed either training field 1 (industrial control sys.) or training field 2 (natural resources);

3) that he very closely, closely, and less closely resembles students who satisfactorily completed training fields 1, 2, and 3 respectively.

In counseling student No. 25, whose choice is engineering (field 3), the vocational counselor may try to make the student aware:

1) that he resembles students who satisfactorily completed training field 3;

2) that he may want to consider either training field 2 or training field 8 which, compared with students who satisfactorily completed these training fields, he resembles closely and less closely respectively.

Student No. 26 resembled students who satisfactorily completed training field 8 and 3, while student No. 27 resembled students who satisfactorily completed training field 2.

From the conditional and Bayesian classifications shown in the print-out, the vocational counselor may counsel incoming students to consider three alternative training fields. However, counselors should not lose sight of the fact that the final decision is to be made by the student.

Implementation of the prediction scheme

It was shown that analysis 2 (graduates) gave fairly good classifications. Also, that was an indication that the group centroids, total dispersion, and transformation matrices, which were derived in this analysis could be used as bases for computing centours (probabilities) and classifying incoming students. However, the input data for each student must contain scores of all 19 predictors.

In the event that the scores of the Alberta Grade 9 departmental examinations and SCAT Level 3 which constituted the 7 of the 19 predictor variables in this study would not be available and since the scores of the 12 predictors were easily accessible to vocational counselors at NAIT, the matrices in analysis 5 could be used instead of those mentioned above. As in the previous case, input data for each student must contain scores for each of the 12 predictors.

The same *a priori* probabilities may be used in computing the classification or perhaps new *a priori* probabilities may be established based on a breakdown of the proposed enrollment of new students for the school term.

Limitations of the study

There were several drawbacks in using data designed for purposes other than this study. Particularly outstanding was the requirement of

a complete set of observations on each student for all predictor variables. Thus, several students had to be dropped from the study due to an incomplete set of scores. However, the sample was large enough to enable the investigator to secure a sufficient number of cases to work with.

Another weakness of this study was that group sample sizes were quite small. The investigator combined several training courses in order to secure adequate group sizes. Also, the sizes of the validation samples were too small to provide a rigorous validation of the results. However, in spite of this weakness, the study showed that the predictors could discriminate fairly well between different training groups.

The choice of predictor variables was also another weakness of this study. Some variables, as shown in the analyses, were somewhat duplicated by other variables and thus contributed little to prediction. Ten predictors out of 19 appeared to contribute greatly to the discrimination of different groups.

However, this would not imply that only a few predictors were needed in this type of prediction scheme. Instead, from the interpretation on how the scores of the sample of graduates were distributed in the reduced discriminant space, it indicated that the use of a large number of predictors would improve differentiation of groups. What was important was the selection of appropriate predictors to measure the various psychological domains of a given population under study. Furthermore, this study showed that 19 predictors differentiated and classified better than 10 or 12 predictors.

Implications for research

This study could be a start of a series of studies along this line at NAIT, as well as in other technological institutes. Validation studies may be conducted using new samples to verify the usefulness of this prediction scheme.

A 2-year experiment might be designed by taking two groups of new students. One group could consist of students who choose their technologies by the usual procedures prescribed at NAIT while the other group could consist of students who choose their technologies through the assistance of a vocational counselor who uses classification based on the prediction scheme of this study. Two years later, comparisons of outcomes could be made to find out which procedure of choosing technologies would result in a large proportion of graduates.

A validation study could be conducted on a new sample of graduates using the data derived in this study and verifying the results of this study. Other predictors, such as the Strong Vocational Interest Blank, the Kuder Preference Record could perhaps be used with the highly weighted variables found in this study and extending the criteria not only to training outcomes but also job success after leaving school. Also, research might be carried out to develop or locate measures which could distinctly discriminate between different training fields. The development of procedures for determining job families still needs further investigation.

Conclusions

1. The null hypothesis that there were no significant differences among group means in the sub-scores of the Differential Aptitude Tests,

the Lorge-Thorndike Intelligence Tests, the Grade Nine Alberta Departmental Examinations, SCAT Level 3, and high school averages of groups of NAIT students, classified on an a priori basis, was rejected. There were differences in group means among the training groups. This indicated that the variables in this study were useful in differentiating various groups.

2. The generalized multivariate null hypothesis that characteristics of various training groups, as measured by the nineteen predictors, were the same, was not tenable. Group centroids were found to be significantly different. This further indicated that the variables could be effective predictors for identifying membership in certain training groups of students enrolled at NAIT.

3. Three roots were found significant and were used in determining the dimensionality of the discriminant space of the graduate sample. It was, therefore, possible to describe the separation of groups along the discriminant functions in these reduced spaces.

4. Results of the classification solutions showed that identification of membership of persons to certain training groups gave a fairly good percentage of correct predictions well beyond the probability of 1 in 8 or .125.

5. Psychological meanings were derived from the distribution of scores of the NAIT graduate criterion sample in a three-dimensional discriminant space. The first discriminant function was interpreted as measuring mechanical characteristics of the students, the second function as measuring intellectual characteristics, and the third as measuring verbal-enterprising characteristics.

6. The discriminant analysis and classification involving 2

outcome categories and 19 predictors showed that it was not possible to differentiate with sufficient accuracy between graduates and non-graduates. However, the results of the analysis involving 8 training groups and 19 predictors would likely be helpful in counseling the student's choice of training program to pursue and could supplement the selection procedures of the institute.

In conclusion, the main objective was to develop a procedure for predicting training outcomes of graduates at NAIT, based on 19 predictor variables. This study demonstrated that the 19 predictors may be analyzed to provide a workable prediction scheme for purposes of vocational counseling. Thus, it is possible to develop prediction schemes for use in technological institutes as well as in high schools.

BIBLIOGRAPHY

Baggaley, A. R. The relation between scores obtained by Harvard freshmen on the Kuder Preference Record and their fields of concentration. Journal of Educational Psychology, 1947, 38, 421-427.

Beaton, A. E. The use of special matrix operators in statistical calculus. (Res. Bull. 64-51) Princeton, N. J.: Educational Testing Service, 1965.

Bennett, G. K., Seashore, H. G., & Wesman, A. G. Manual for the Differential Aptitude Tests. (3rd ed.) New York: Psychological Corporation, 1959.

Borrow, H. (Ed.) Man in a world at work. Boston: Houghton Mifflin, 1964.

Bryan, J. G. A method for the exact determination of the characteristic equation and the latent vectors of a matrix with application to the discriminant function for more than two groups. Unpublished doctoral dissertation, Harvard Graduate School of Education, 1950.

Bryan, J. G. The generalized discriminant function: Mathematical foundation and computational routine. Harvard Educational Review, 1951, 21(1), 90-95.

Calia, V. F. The use of discriminant analysis in the prediction of scholastic performance. Personnel and Guidance Journal, 1960, 39, 184-192.

Cass, J. C., & Tiedeman, D. V. Vocational development and the election of a high school curriculum. Personnel and Guidance Journal, 1960.

Cooley, W. W. The application of a developmental rationale and methods of multivariate analysis of the study of potential scientists. Unpublished doctoral dissertation, Harvard Graduate School of Education, 1958.

Cooley, W. W. A computer-measurement system for guidance. Harvard Educational Review, 1964, 34(4), 559-572.

Cooley, W. W., & Lohnes, P. R. Multivariate procedures for the behavioral sciences. New York: Wiley, 1962.

Cronbach, L. J. Essentials of psychological testing. (2nd ed.) New York: Harper & Row, 1960.

Cronbach, L. J., & Gleser, G. C. Psychological tests and personnel decisions. Urbana: University of Illinois Press, 1965.

Cureton, E. E. A note on the use of Burt's formula for estimating factor significance. The British Journal of Psychology, Statistical Section, 1955, 8, 28.

Division of Educational Research Services (D.E.R.S.). Computer Program Documentation. Edmonton: University of Alberta, 1968.

Dunn, F. E. Two methods for predicting the selection of a college major. Journal of Counseling Psychology, 1959, 6(1), 15-21.

Fisher, R. A., & Yates, F. Statistical tables for biological agricultural and medical research. (4th ed.) Edinburgh: Oliver and Boyd, 1953.

French, W. L. Can a man's occupation be predicted? Journal of Counseling Psychology, 1959, 6(2), 95-101.

Garret, H. E. The discriminant function and its use in psychology. Psychometrika, 1943, 8(2), 65-79.

Goldman, L. Using tests in counseling. New York: Appleton-Century-Crofts, 1961.

Guilford, J. P. Personality. New York: McGraw-Hill, 1959.

Guilford, J. P., Hoepfner, R., & Petersen, H. Predicting achievement in ninth-grade mathematics from measures of intellectual-aptitude factors. Educational and Psychological Measurement, 1965, 25(3), 659-682.

Horst, P. Psychological measurement and prediction. Belmont, California: Wadsworth Publishing Company, Inc., 1966.

King, R. G. The prediction of undergraduate field of concentration in Harvard College. Unpublished doctoral dissertation, Harvard Graduate School of Education, 1958.

Lohnes, P. R. Test space and discriminant space classification models and related significance tests. Educational and Psychological Measurement, 1961, 21(3), 559-573.

Lorge, I., & Thorndike, R. L. Technical manual for Lorge-Thorndike Intelligence Tests. Boston: Houghton-Mifflin, 1954.

Madaus, G. F., & O'Hara, R. P. Vocational interest patterns of high school boys: A multivariate approach. Journal of Counseling Psychology, 1967, 14(2), 106-112.

Pearson, K. Tables for statisticians and biometricians. Vol. 1 (2nd ed.) University College, London: The Biometric Laboratory, 1924.

Rao, C. R. Advanced statistical methods in biometric research. New York: Wiley, 1952.

Rao, C. R. Linear statistical inference and its applications. New York: Wiley, 1965.

Rulon, P. J. Distinctions between discriminant and regression analysis and a geometric interpretation of the discriminant function. Harvard Educational Review, 1951, 21(1), 80-90.

Rulon, P. J., Tiedeman, D. V., Tatsuoka, M. M., & Langmuir, C. R. Multivariate statistics for personnel classification. New York: Wiley, 1967.

Stahmann, R. F., & Wallen, N. E. Multiple discriminant prediction of major field of study. Educational and Psychological Measurement, 1966, 26(2), 439-444.

Stinson, P. J. A method for counseling engineering students. Personnel and Guidance Journal, 1958, 37, 294-295.

Strong, E. K., Jr. Vocational interests 18 years after college. Minneapolis: University Press, 1955.

Super, D. E. Guidance: Manpower utilization or human development? Personnel and Guidance Journal, 1954, 33, 8-14.

Thorndike, R. L. Personnel selection. New York: Wiley, 1949.

Thorndike, R. L., & Hagen, E. Ten thousand careers. New York: Wiley, 1959.

Tiedeman, D. V. The utility of the discriminant function in psychological and guidance investigations. Harvard Educational Review, 1951, 21(1), 71-79.

Tiedeman, D. V. A model for the profile problem. In Anastasi, A. (Ed.), Testing problem in perspective. Washington, D.C.: American Council of Education, 1966.

Tiedeman, D. V., & Bryan, J. G. Prediction of college field of concentration. Harvard Educational Review, 1954, 24, 122-139.

Tiedeman, D. V., Bryan, J. G., & Rulon, P. J. The utility of the airman classification battery for assignment for airmen to eight air force specialities. (Reprinted: Educational Research Corporation, March, 1953). Cambridge, Mass: Educational Research Corporation, 1951.

Tiedeman, D. V., & Sternberg, J. J. Information appropriate for curriculum guidance. Harvard Educational Review, 1952, 22(4), 257-274.

Travers, R. M. W. The use of a discriminant function in the treatment of psychological group differences. Psychometrika, 1939, 4(1), 25-32.

Tyler, L. E. Toward a workable psychology of individuality. American Psychologist, 1959, 14(2), 75-81.

Vacchiano, R. B., & Adrian, R. J. Multiple discriminant prediction of college career choice. Educational and Psychological Measurement, 1966, 26, 985-995.

Wert, J. E., Neidt, C. O., & Ahmann, J. S. Statistical methods in educational and psychological research. New York: Appleton-Century-Crofts, 1954.

APPENDIX I

Mathematical concept of multiple discriminant analysis

The general description of the multiple discriminant analysis as described by Tiedeman, Byran, and Rulon (1953) is quoted in full below:

Let X_{pgj} denote the value of the j^{th} variate ($j = 1, 2, \ldots, n$) of the p^{th} individual ($p = 1, 2, \ldots, N_g$) in the g^{th} group ($g = 1, 2, \ldots, G$) where G is less than n. Without loss of generality we may define the X_{pgj} so that

$$\sum_g \sum_p X_{pgj} = 0 \qquad (2.1)$$

for all values of j. In general, separate group means such as

$$\bar{X}_{gj} = \frac{1}{N_g} \sum_p X_{pgj} \qquad (2.2)$$

do not vanish.

A linear function of the X_j such as

$$y_g = v_1X_1 + v_2X_2 + \ldots + v_nX_n \qquad (2.3)$$

would have group means of

$$y_g = \frac{1}{N_g} \sum_p y_{pg} \qquad (2.4)$$

where y_{pg} is the value of the linear function for the p^{th} individual in the g^{th} group. Because of the restrictions defined by equation (2.1), the among means of groups sum of squares of y is

$$\sum_g N_g \bar{y}_g^2 \qquad (2.5)$$

The within groups sum of squares is

$$\sum_g \sum_p (y_{pg} - \bar{y})^2 \tag{2.6}$$

The coeficients, $v_1, v_2, \ldots, v_n$ of y are defined by Fisher's[3] criterion of discriminant analysis, i.e., by maximization of the ratio

$$\lambda = \frac{\sum_g N_g \bar{y}_g^2}{\sum_g \sum_p (y_{pg} - \bar{y})^2} \tag{2.7}$$

The function λ, has several extrema each of which is indicative of a distinct dimension of the subspace defined by the group means. All discriminant functions are obtained from the same initial ratio λ.[4]

Maximization of λ is accomplished by a combination of two effects, increase in the among means of groups sum of squares, and decrease in the within groups sum of squares. Therefore, the linear functions defined by the discriminant analysis criterion provide as much separation of group centroids as is possible within the restriction of having the groups as homogeneous as possible. The utility for group classification of functions with this property is obvious.

Manipulation of the quadratic forms of the variates required for discriminant analysis is facilitated by use of matrix algebra. Therefor let us define the symmetrical matrices

$$A = \left\| a_{ij} \right\|, \quad a_{ij} = \sum_g N_g \bar{X}_{gi} \bar{X}_{gj} = a_{ji}, \quad (i, j = 1, 2, \ldots, n); \tag{2.8}$$

---

[3] R.A. Fisher, "The Use of Multiple Measurement in Taxonomic Problems," Annals of Eugenics, VII (1936) 179-188.

[4] J. G. Bryan, A Method for the Exact Determination of the Characteristic Equation of a Matrix with Applications to the Discriminant Function for More than Two Groups. Cambridge, Mass.: Harvard Graduate School of Education, 1950.

$$W = \left\| w_{ij} \right\| , \; w_{ij} = \sum_g \sum_p (X_{pgi} - \bar{X}_{gi}) (X_{pgj} - \bar{X}_{gj}) = w_{ji}, \; (i, j = 1, 2, \ldots, n); \tag{2.9}$$

and the column vector

$$v = [v_j] , \; (j = 1, 2, \ldots, n). \tag{2.10}$$

In this notation, equations (2.5), (2.6), and (2.7) may be written as

$$\sum_g N_g \bar{y}_g^2 = v'Av \tag{2.5}$$

$$\sum_g \sum_p (y_{pg} - \bar{y}_g)^2 = v'Wv \tag{2.6}$$

and $\lambda = \frac{v'Av}{v'Wv}$ (2.7)

It has been shown[5] that the matrix equation defining the vector v is

$$(v'Wv)\, Av - (v'Av)\, Wv = 0$$

This may be rewritten as

$$(A - \lambda W)\, v = 0$$

Defining

$$R \equiv W^{-1} A,$$

this becomes

$$(R - \lambda I)v = 0 \tag{2.11}$$

where I is the unit matrix.

The coefficient of the discriminant functions are determined by the latent vectors of R, and the corresponding latent roots of R equal the respective ratios of among-groups to within-groups sums of squares. By considering the rank of the matrix A, it is a simple matter to show that the number of solutions of (2.11), such that $\lambda \neq 0$, is at most

---

[5] See for instance: J. G. Bryan, ibid., pp. 131-132.

equal to the smaller of the two integers G - 1, and n. Consequently, letting r stand for the smaller number, the total discriminative power of the variates is exhausted by r linear functions defined in the manner stated. Among these, all functions corresponding to distinct values of λ are uncorrelated as they stand. Repeated roots other than zero are possible but unlikely to occur. If one or more multiple roots should occur, however, the vectors corresponding to any one of them are already uncorrelated with the vectors corresponding to all different roots and can be chosen in such a way as to be uncorrelated among themselves. The numerical values of these functions are independent of the origin of coordinates, the units of measurement, and in fact independent of any non-singular linear transformation of the variates.[6]

---

[6]J. G. Bryan, ibid., pp. 138-139.

```
LEVEL 16 ( 1 JULY 68)                    OS/360  FORTRAN H                          DATE  69.172/05.04.20

          COMPILER OPTIONS - NAME=  MAIN,OPT=02,LINECNT=59,SOURCE,EBCDIC,NOLIST,NODECK,LOAD,NOMAP,NOEDIT,ID,NOXREF
                C    MULV10          DIVISION OF EDUCATIONAL RESEARCH SERVICES
                C                    UNIVERSITY OF ALBERTA
                C         ************************************************************
                C    PURPOSE:        CARRIES OUT DISCRIMINANT ANALYSIS AND ONE WAY
                C                    MULTIVARIATE ANALYSIS OF VARIANCE, AND ALSO GIVES CARD
                C                    OUTPUT FOR CLASSIFYING SUBJECTS BY USING MULVD3
                C    CARD INPUT:
                C          1.TITLE(20A4)
                C          2.PARAMETERS(16I5):NG,NV,NT,ND,(NCELL(I),I=1,NG)
                C          3.FORMAT FOR DATA CARDS(20A4)
                C          4.DATA CARDS
                C          5.A BLANK CARD
                C    DESCRIPTION OF PARAMETERS:
                C              NG: NO OF GROUPS
                C              NV: NO OF VARIABLES PER OBSERVATION
                C              NT: NO OF TOTAL OBSERVATIONS
                C              ND: TYPE 1 IF CARD OUT PUT FOR MULV11 REQUIRED.
                C              NCELL(I): NO OF OBSERVATIONS IN I TH GROUP
                C    REMARKS:
                C          1.NV AND NG ARE LIMITED UP TO 50
                C          2.EACH GROUP SHOULD HAVE AT LEAST NV OBSERVATIONS I.E.,
                C            NCELL(I) SHOULD NOT BE LESS THAN NV
                C    PROGRAMMER      K.BAY
                C    SUPPROGRAMS    FISHER,HOMODIS,MXOUT,DATA,RAOFT,ROYTES,DATRAN,SIGDIS
                C              SSP CORRE,GMADD,GMPRO,GMSUB,MCPY,MINV,CSUM,LOC,NROOT,EIGEN
                C
  ISN 0002            DIMENSION FMT(20),TITLE(20),X(50),NCELL(50),XBAR(50),STD(50),
                     1BB(50),DD(50),TT(50),LL(50),MM(50),COR(1300)
  ISN 0003            DIMENSION R(50,50),W(2500),H(2500),E(50,50),D(2500),RX(2500)
  ISN 0004        100 FORMAT(20A4)
  ISN 0005        101 FORMAT(1H1,10X,20A4)
  ISN 0006        102 FORMAT(16I5)
  ISN 0007        103 FORMAT(1X,'NO. OF GROUPS =',22X,I5,/,1H0,'NO OF VARIABLES PER OBSE
                     1RVATION =',4X,I5,/,1H0,'TOTAL  NO OF OBSERVATIONS =',10X,I5)
  ISN 0008        104 FORMAT(1H0,'FORMAT FOR DATA :',20A4)
  ISN 0009        105 FORMAT(/,/,1H0,'GROUP:',I3,4X,'NO OF OBSERVATION IN THIS GROUP',I5
                     1)
  ISN 0010        106 FORMAT(1H1,'BARTLETTS HOMOGENITY OF DISPERSION TEST SEE COOLEY  P
                     1 162')
  ISN 0011        107 FORMAT(1H1,'WILKS LAMDA TEST SEE COOLEY P 61')
  ISN 0012        108 FORMAT(1X,'DETERMINANT OF W=',F16.8)
  ISN 0013        109 FORMAT(1X,'DF1=',F5.0,3X,'DF2=',F9.0,3X,'F-RATIO =',E14.6,3X,
                     1'PROBABILITY=',F9.6,3X,'LAMDA=',F9.6)
  ISN 0014        110 FORMAT(1H1,'MAXMUM LATENT ROOT APPROACH SEE MORRISON P168')
  ISN 0015        111 FORMAT(1H0,'S =',I5,3X,'M =',F9.1,3X,'N =',F9.1,3X,'HECK =',F9.6)
  ISN 0016        112 FORMAT(1H0,' CONSULT HECKS CHART ON MORRISON P 312')
  ISN 0017        113 FORMAT(5F16.9)
  ISN 0018        114 FORMAT(1H1,'DISCRIMINANT ANALYSIS')
  ISN 0019            MAX=50
  ISN 0020            MAXM=MAX*MAX
  ISN 0021          1 READ(5,100,END=9999) TITLE
  ISN 0022            IF (TITLE(1).EQ.TITLE(2)) GO TO 9999
  ISN 0024            WRITE(6,101) TITLE
  ISN 0025            READ(5,102) NG,NV,NT,ND,(NCELL(I),I=1,NG)
  ISN 0026            WRITE(6,103) NG,NV,NT
  ISN 0027            READ (5,100)FMT
  ISN 0028            WRITE(6,104) FMT
  ISN 0029            DO 2 I=1,MAXM
  ISN 0030          2 W(I)=0.0
  ISN 0031            HILOGS=0.0
  ISN 0032            REWIND 2
  ISN 0033            DO 3 J=1,NT
  ISN 0034            READ(5,FMT) (X(I),I=1,NV)
  ISN 0035            CALL DATRAN (NV,X)
  ISN 0036          3 WRITE(2) (X(I),I=1,NV)
  ISN 0037            REWIND 2
  ISN 0038            NVV=NV*NV
  ISN 0039            DO 5 NNG=1,NG
  ISN 0040            NN=NCELL(NNG)
  ISN 0041            WRITE(6,105) NNG,NN
  ISN 0042            CALL CORRE (NN,NV,0,0.0,XBAR,STD,RX,COR,BB,DD,TT)
  ISN 0043            IF(ND.EQ.1) WRITE(7,113) (XBAR(I),I=1,NV)
  ISN 0045            CALL MXOUT(XBAR,1,NV,0,60,132,1,24,24H MEANS FOR THIS GROUP    )
  ISN 0046            CALL MXOUT(STD,1,NV,0,60,132,1,36,36H STANDARD DEVIATIONS FOR THIS
                     1 GROUP )
  ISN 0047            CALL GMADD(W,RX,W,NV,NV)
  ISN 0048            DO 4 I=1,NVV
  ISN 0049          4 RX(I)=RX(I)/(NN-1.0)
  ISN 0050            CALL MXOUT(RX,NV,NV,0,60,132,1,20,20H COVARIANCE MATRIX    )
  ISN 0051            IF(ND.NE.1) GO TO 15
  ISN 0053            CALL ARRAY(1,NV,NV,50,50,RX,E)
  ISN 0054            DO 14 I=1,NV
  ISN 0055         14 WRITE(7,113) (E(I,J),J=1,NV)
  ISN 0056         15 CALL MINV(RX,NV,DET,LL,MM)
  ISN 0057            DETERM=ALOG (DET)
  ISN 0058            HILOG=( NN-1.0)*DETERM
  ISN 0059            HILOGS=HILOGS+HILOG
  ISN 0060          5 CONTINUE
  ISN 0061            REWIND 2
  ISN 0062            CALL MCPY(W,E,NV,NV,0)
  ISN 0063            CALL MINV(E,NV,DETG,LL,MM)
  ISN 0064            CALL CORRE(NT,NV,0,0.0,XBAR,STD,RX,COR,BB,DD,TT)
  ISN 0065            WRITE(6,101)
  ISN 0066            CALL GMSUB(RX,W,H,NV,NV)
  ISN 0067            CALL MXOUT(XBAR,1,NV,0,60,132,1,24,24H MEANS OF TOTAL SAMPLE   )
  ISN 0068            CALL MXOUT(STD,1,NV,0,60,132,1,40,40H STANDARD DEVIATIONS FOR TOTA
                     1L SAMPLE       )
  ISN 0069            CALL MXOUT(COR,NV,NV,1,60,132,1,36,36H CORRELATION MATRIX FOR TOTA
                     1L SAMPLE)
  ISN 0070            DO 6 I=1,NVV
  ISN 0071          6 D(I)=W(I)/(NT-NG)
  ISN 0072            CALL MINV(D,NV,DET,LL,MM)
  ISN 0073            DETERM=ALOG(DET)
  ISN 0074            WRITE(6,106)
  ISN 0075            WRITE(6,108) DET
```

```
ISN 0076            HILOG=(NT- NG)*DETERM
ISN 0077            XMM=HILOG-HILOGS
ISN 0078            CALL MINV(RX,NV,OFT,LL,MM)
ISN 0079            OETERT=ALOG(DFT)
ISN 0080            WRITF(6,107)
ISN 0081            NDF=NG-1
ISN 0082            CALL RAOFT(H,W,R,LL,MM,NV,NOF,DETG,XLAMB,NG,NT,F1,F2,F,PRO)
ISN 0083            WRITF(6,109) F1,F2,F,PRO,XLAMB
ISN 0084            REWIND 2
ISN 0085            WRITE(6,110)
ISN 0086            CALL MXOUT(W,NV,NV,0,60,132,1,12,12H E MATRIX   )
ISN 0087            CALL MXOUT(H,NV,NV,0,60,132,1,12,12H H MATRIX   )
ISN 0088            CALL GMPRD(H,E,R,NV,NV,NV)
ISN 0089            CALL MXOUT(R,NV,NV,0,60,132,I,12,12H HE-1 MATRIX)
ISN 0090            CALL MCPY(W,E,NV,NV,0)
ISN 0091            CALL MCPY(W,D,NV,NV,0)
ISN 0092            CALL ROYTES(W,H,R,X,NOF,NV,NG,NT,NS,FM,FN,RDDT)
ISN 0093            WRITE(6,111) NS,FM,FN,ROOT
ISN 0094            WRITE(6,112)
ISN 0095            IF (NO.NE.1) GO TO 8
ISN 0097            CALL ARRAY(1,NV,NV,50,50,E,E)
ISN 0098            DO 7 I=1,NV
ISN 0099            DO 11 J=1,NV
ISN 0100         11 E(I,J)=E(I,J)/(NT-NG)
ISN 0101          7 WRITE(7,113) (E(I,J),J=1,NV)
ISN 0102          8 WPITE(6,114)
ISN 0103            CALL MXOUT(R,NV,NOF,0,60,132,1,40,40H NORMALIZED WEIGHTS FOR CLASS
                   1IFICATION  )
ISN 0104            DO 12 I=1,NVV
ISN 0105         12 D(I)=O(I)/(NT-NG)
ISN 0106            CALL MXDUT(O,NV,NV,0,60,132,1,20,20H DISPERSION MATRIX  )
ISN 0107            IF (NO.NF.I) GO TO 10
ISN 0109            CALL ARRAY(1,NV,NOF,50,50,R,R)
ISN 0110            OO 9 I=1,NV
ISN 0111          9 WRITE(7,113) (R(I,J),J=1,NOF)
ISN 0112         10 CALL SIGOS(X,NV,NOF,NG,NT)
ISN 0113            GO TO 1
ISN 0114       9999 STOP
ISN 0115            ENO

****** END OF COMPILATION ******
LEVEL 16 ( I JULY 68)                         OS/360  FORTRAN H                         DATF  69.172/05.04.34

          COMPILER OPTIONS - NAME=  MAIN,OPT=02,LINECNT=59,SOURCE,EBCOIC,NOLIST,NOOECK,LUAO,NOMAP,NOEOIT,IO,NOXREF
ISN 0002            FUNCTION FISHER(FM,FN,X)
              C     COMPUTES PROBABILITY OF F-STATISTIC X WITH M ANO N OEGREES OF
              C     FREEDOM IN NUMERATOR AND DENOMINATOR RESPECTIVELY.
              C     TRANSLATED FROM ALGOL TO 360 FORTRAN IV BY FLATHMAN, APRIL, 1968.
              C     ALGORITHM 322, COMMUNICATIONS OF THE ACM 11,2,115, FEBRUARY,1968.
ISN 0003            M=FM+0.5
ISN 0004            N=FN+0.5
ISN 0005            FISHEP=1.0
ISN 0006            IF(X.LE.0..OR.M.LE.O.OR.N.LE.O) RETURN
ISN 0008            INTEGER A,B
ISN 0009            A=2*(M/2)-M+2
ISN 0010            B=2*(N/2)-N+2
ISN 0011            W=X*M/N
ISN 0012            Z=1./(1.+W)
ISN 0013            IF(A.NE.1)GO TO 2
ISN 0015            IF(B.NE.1)GO TO 1
ISN 0017            P=SQRT(W)
ISN 0018            Y=0.3183099
ISN 0019            D=Y*Z/P
ISN 0020            P=2.*Y*ATAN(P)
ISN 0021            GO TO 4
ISN 0022      1     P=SQRT(W*Z)
ISN 0023            D=0.5*P*Z/W
ISN 0024            GO TO 4
ISN 0025      2     IF(B.NE.1)GO TO 3
ISN 0027            P=SQRT(Z)
ISN 0028            D=0.5*Z*P
ISN 0029            P=1.-P
ISN 0030            GO TO 4
ISN 0031      3     D=Z*Z
ISN 0032            P=W*Z
ISN 0033      4     Y=2.*W/Z
ISN 0034            J=B+2
ISN 0035            IF(J.GT.N)GO TO 6
ISN 0037      5     D=(1.+A/(J-2.))*O*Z
ISN 0038            IF(A.EQ.1) P=P+D*Y/(J-1)
ISN 0040            IF(A.NE.1) P=(P+W)*Z
ISN 0042            J=J+2
ISN 0043            IF(J.LE.N) GO TO 5
ISN 0045      6     Y=W*Z
ISN 0046            Z=2./Z
ISN 0047            B=N-2
ISN 0048            I=A+2
ISN 0049            IF(I.GT.M) GO TO 8
ISN 0051      7     J=I+B
ISN 0052            D=Y*O*J/(I-2)
ISN 0053            P=P-Z*O/J
ISN 0054            I=I+2
ISN 0055            IF(I.LE.M) GO TO 7
ISN 0057      B     FISHER=1.-P
ISN 0058            RETURN
ISN 0059            END

****** END OF COMPILATION ******
LEVEL 16 ( 1 JULY 68)                         OS/360  FORTRAN H                         DATE  69.172/05.04.40

          COMPILER OPTIONS - NAME=  MAIN,OPT=02,LINECNT=59,SOURCE,EBCDIC,NOLIST,NODECK,LOAO,NOMAP,NOEDIT,ID,NOXREF
ISN 0002            SUBROUTINE HOMDIS(XMM,NCELL,NV,NG,NT)
              C   USING BARTLETTS METHOD, TEST HOMOGENITY OF OISPERSION MATRICES
              C     XMM      INPUT ((NT-NG)*ALOG DET POOLEO W)-SUM OF ((NCELL(I)-1)*
              C              ALOG DET W(I))
              C     NCELL    NO OF OBSERVATION IN EACH SAMPLE
              C     NV       SIZE OF OISPERSION MATRICES
              C     NG       LENGTH OF NCELL
```

```
                 C      NT       TOTAL NO OF OBSERVATIONS.
   ISN 0003             DIMENSION NCELL(1)
   ISN 0004         100 FORMAT(1X,'DF1=',F9.1,2X,'DF2=',F9.1,2X,'F-RATIO=',E16.8,2X,
                       1'PROBABILITY=',F9.6)
   ISN 0005         111 FORMAT(1X,'FOR TEST OF H1 M= ',E16.8)
   ISN 0006         113 FORMAT(1X,'A2-A1 SQUARED =',E16.8)
   ISN 0007         116 FORMAT(1X,'A1 =',F14.7)
   ISN 0008         117 FORMAT(1X,'A2 =',F14.7)
   ISN 0009         118 FORMAT(1X,'B=',E16.8)
   ISN 0010             FAIS=0.0
   ISN 0011             GAIS=0.0
   ISN 0012             DO 1 I=1,NG
   ISN 0013             FAIS=FAIS+1.0/(NCELL(I)-1.0)
   ISN 0014           1 GAIS=GAIS+1.0/((NCELL(I)-1.0)**2)
   ISN 0015             F1=.5*( NG-1.0)* NV*( NV+1.0)
   ISN 0016             A1A=(FAIS-(1.0/(NT- NG)))*((2.0*( NV**2.0))+(3.0* NV)-1.0)
   ISN 0017             A1=A1A/(6.0*( NG-1.0)*( NV+1.0))
   ISN 0018             A2=(GAIS-(1.0/(NT- NG)**2))*(( NV-1.0)*( NV+2.0))/(6.0*( NG-1.0))
   ISN 0019             DIF=A2-(A1**2.0)
   ISN 0020             WRITE(6,111) XMM
   ISN 0021             WRITE(6,113) DIF
   ISN 0022             IF (DIF) 8,8,9
   ISN 0023           8 F2=(F1+2.0)/((A1**2.0)-A2)
   ISN 0024             B=F2/(1.0-A1+(2.0/F2))
   ISN 0025             F=(F2*XMM)/(F1*(B-XMM))
   ISN 0026             GO TO 10
   ISN 0027           9 F2=(F1+2.0)/DIF
   ISN 0028             B=F1/(1.0-A1-(F1/F2))
   ISN 0029             F =XMM/B
                 C F IS FRATIO FOR TEST OF HOMOGENITY OF DISPERSION WITH F1,F2 DF
   ISN 0030          10 PRO=FISHER(F1,F2,F)
   ISN 0031             WRITE(6,100) F1,F2,F,PRO
   ISN 0032             WRITE(6,116)     A1
   ISN 0033             WRITE(6,117) A2
   ISN 0034             WRITE(6,118) B
   ISN 0035             RETURN
   ISN 0036             END

****** END OF COMPILATION ******
LEVEL 16 ( 1 JULY 68)                           OS/360  FORTRAN H                            DATE  69.172/05.04.46

          COMPILER OPTIONS - NAME=  MAIN,OPT=02,LINECNT=59,SOURCE,EBCDIC,NOLIST,NODECK,LOAD,NOMAP,NOEDIT,ID,NOXREF
   ISN 0002             SUBROUTINE MXOUT(A,N,M,MS,LINS,IPOS,ISP,NUMHOL,TITLE)
                 C
                 C        A       :NAME OF OUTPUT MATRIX
                 C        N       :NUMBER OF ROWS IN A
                 C        M       :NUMBER OF COLUMNS IN A
                 C        MS      :STORAGE MODE OF A
                 C                   0 GENERAL
                 C                   1 SYMETRIC
                 C                   2 DIAGONAL
                 C        LINS    :NUMBER OF PRINT LINES ON THE PAGE(USUALLY 60)
                 C        IPOS    :NUMBER OF PRINT POSITIONS ACROSS THE PAGE(USUALLY 132)
                 C        ISP     :LINE SPACING CODE, 1 FOR SINGLE SPACE 2 FOR DOUBLE
                 C                 SPACE
   ISN 0003             DIMENSION A(1),B(8),TITLE(20)
   ISN 0004           1  FORMAT (1H0,/,1H0,20A4)
   ISN 0005           2 FORMAT(12X,8HCOLUMN   ,7(3X,I3,10X))
   ISN 0006           3 FORMAT(1H )
   ISN 0007           4 FORMAT(1H ,7X,4HROW ,I3,7(E16.6))
   ISN 0008           5 FORMAT(1H0,7X,4HROW ,I3,7(E16.6))
   ISN 0009             NN=(NUMHOL+3)/4
   ISN 0010             WRITE(6,1) (TITLE(J),J=1,NN)
   ISN 0011             J=1
   ISN 0012             NEND=IPOS/16-1
   ISN 0013             LEND=(LINS/ISP)-2
   ISN 0014          10 LSTRT=1
   ISN 0015          20 CONTINUE
   ISN 0016             JNT=J+NEND-1
   ISN 0017          31 IF(JNT-M)33,33,32
   ISN 0018          32 JNT=M
   ISN 0019          33 CONTINUE
   ISN 0020             WRITE(6,2) (JCUR,JCUR=J,JNT)
   ISN 0021             IF(ISP-1) 35,35,40
   ISN 0022          35 WRITE(6,3)
   ISN 0023          40 LTEND=LSTRT+LEND-1
   ISN 0024             DO 80 L=LSTRT,LTEND
   ISN 0025             DO 55 K=1,NEND
   ISN 0026             KK=K
   ISN 0027             JT=J+K-1
   ISN 0028             CALL LOC(L,JT,IJNT,N,M,MS)
   ISN 0029             B(K)=0.0
   ISN 0030             IF(IJNT) 50,50,45
   ISN 0031          45 B(K)=A(IJNT)
   ISN 0032          50 CONTINUE
   ISN 0033             IF(JT-M) 55,60,60
   ISN 0034          55 CONTINUE
   ISN 0035          60 IF(ISP-1) 65,65,70
   ISN 0036          65 WRITE(6,4) L,(B(JW),JW=1,KK)
   ISN 0037             GO TO 75
   ISN 0038          70 WRITE(6,5) L,(B(JW),JW=1,KK)
   ISN 0039          75 IF(N-L) 85,85,80
   ISN 0040          80 CONTINUE
   ISN 0041             LSTRT=LSTRT+LEND
   ISN 0042             GO TO 20
   ISN 0043          85 IF(JT-M) 90,95,95
   ISN 0044          90 J=JT+1
   ISN 0045             GO TO 10
   ISN 0046          95 RETURN
   ISN 0047             END

****** END OF COMPILATION ******
LEVEL 16 ( 1 JULY 68)                           OS/360  FORTRAN H                            DATE  69.172/05.04.54

          COMPILER OPTIONS - NAME=  MAIN,OPT=02,LINECNT=59,SOURCE,EBCDIC,NOLIST,NODECK,LOAD,NOMAP,NOEDIT,ID,NOXREF
   ISN 0002             SUBROUTINE DATA(N,X)
   ISN 0003             DIMENSION X(1)
   ISN 0004             READ (2) (X(I),I=1,N)
```

```
   ISN 0005              RETURN
   ISN 0006              END

****** END OF COMPILATION ******
LEVEL 16 ( 1 JULY 68)                          OS/360  FORTRAN H                                DATE  69.172/05.04.58

        COMPILER OPTIONS - NAME=  MAIN,OPT=02,LINECNT=59,SOURCE,EBCDIC,NOLIST,NODECK,LOAD,NOMAP,NOEDIT,ID,NOXREF
   ISN 0002              SUBROUTINE RAOFT(SA ,SE,SH,LL,MM,NP,NOF,DET,ALAMDA,NR,NT,F1,F2,FR,
                        1PRO)
                C     PUPOSE PERFORMS RAOS APPROXMATE F TEST
                C       SA              D.P. MATRIX FOR HYPOTHESIS
                C       SE              D.P. MATRIX FOR ERROR
                C       SH              WORKING VECTOR OF LENGTH NP*NP
                C       LL,MM           WORKING VECTOR OF LENGTH NP
                C       NP              SIZE OF MATRIX SA,SE
                C       NDF             RANK OF C MATRIX  NG-1
                C       DET             INPUT DETERMINANT OF SE MATRIX
                C       ALAMDA          OUPUT WILKS LAMDA
                C       NR              RANK OF DESIGN MATRIX OR NG
                C       NT              TOTAL NO OF OBSERVATIONS
                C       F1              OUTPUT DF1
                C       F2              OUTPUT DF2
                C       FR              OUT PUT FRATIO
                C       PRO             OUT PUT PROBABILITY
   ISN 0003              DIMENSION SA(1),SH(1),SE(1),LL(1),MM(1)
   ISN 0004              CALL GMADD(SA,SE,SH,NP,NP)
   ISN 0005              CALL MINV(SH,NP,DA,LL,MM)
   ISN 0006              ALAMDA=DET/DA
   ISN 0007              NS=MINO(NP,NDF)
   ISN 0008              IF(NS.EQ.1) GO TO 10
   ISN 0010              IF(NS.EQ.2) GO TO 11
   ISN 0012              S=SQRT(((NP**2)*(NDF**2)-4.0)/(NP**2+NDF**2-5.0))
   ISN 0013              F1=NP*NDF
   ISN 0014              FM=NT-NR-(NP-NDF+1.0)/2.0
   ISN 0015              TT=(NP*NDF-2.0)/2.0
   ISN 0016              F2=EM*S-TT
   ISN 0017              AL=ALAMDA**(1.0/S)
   ISN 0018              GO TO 20
   ISN 0019           10 F1=1.0+IABS(NDF-NP)
   ISN 0020              F2=1.0+NT-NR-NP
   ISN 0021              AL=ALAMDA
   ISN 0022              GO TO 20
   ISN 0023           11 F1=4.0+2.0*IABS(NDF-NP)
   ISN 0024              F2=2.0+2.0*(NT-NR-NP)
   ISN 0025              AL=SQRT(ALAMDA)
   ISN 0026           20 FR=((1.0-AL)*F2)/(AL*F1)
   ISN 0027              PRO=FISHER(F1,F2,FR)
   ISN 0028              RETURN
   ISN 0029              END

****** END OF COMPILATION ******
LEVEL 16 ( 1 JULY 68)                          OS/360  FORTRAN H                                DATE  69.172/05.05.04

        COMPILER OPTIONS - NAME=  MAIN,OPT=02,LINECNT=59,SOURCE,EBCDIC,NOLIST,NODECK,LOAD,NOMAP,NOEDIT,ID,NOXREF
   ISN 0002              SUBROUTINE ROYTES(SE,SA,SH,XLA,NOF,NP,NR,NT,NS,FM,FN,AMAX)
                C          COMPUTE EIGENVALUES OF (SA)*(SE-1)
                C          SE INPUT MATRIX
                C          SA INPUT MATRIX
                C          SH OUT PUT  EIGENVECTORS
                C          XLA  OUTPUT EIGENVALUES VECTOR
                C          NOF   INPUT DF
                C          NP   INPUT NO OF VARIABLES
                C          NR THE RANK OF DESIGN MATRIX
                C          NT   TOTAL NO OF OBSERVATION
                C          NS FM FN AMAX PARAMETERS FOR MAX ROOT TEST
   ISN 0003              DIMENSION SE(1),SA(1),SH(1),XLA(1)
   ISN 0004              CALL NROOT(NP,SA,SE,XLA,SH)
   ISN 0005              CALL MXOUT(XLA,1,NP,0,60,132,1,12,12H EIGENVALUES)
   ISN 0006              AMAX=-999999.0
   ISN 0007              DO 1 I=1,NP
   ISN 0008            1 IF(XLA(I).GT.AMAX) AMAX=XLA(I)
   ISN 0010              AMAX=AMAX/(1.0+AMAX)
   ISN 0011              NS=MINO(NOF,NP)
   ISN 0012              FM=(IABS(NOF-NP)-1.0)/2.0
   ISN 0013              FN=(NT-NR-NP-1.0)/2.0
   ISN 0014              RETURN
   ISN 0015              END

****** END OF COMPILATION ******
LEVEL 16 ( 1 JULY 68)                          OS/360  FORTRAN H                                DATE  69.172/05.05.09

        COMPILER OPTIONS - NAME=  MAIN,OPT=02,LINECNT=59,SOURCE,EBCDIC,NOLIST,NODECK,LOAD,NOMAP,NOEDIT,ID,NOXREF
   ISN 0002              FUNCTION CHIPRB(CHI,NDF)
                C        COMPUTES PROBABILITY OF CHISQUARE CHI WITH NDF DEGREES OF FREEDOM.
   ISN 0003              EXTERNAL ERF,SQRT
   ISN 0004              REAL NORMAL
   ISN 0005              INTEGER F
   ISN 0006              LOGICAL BIGX,EVEN
   ISN 0007              NORMAL(X)=0.5*(1.0+ERF(0.7071068*X))
                C        SEE KENNEY AND KEEPING(1951) VOL.2, P.43, WHERE ERF IS CALLED G.
   ISN 0008              F=NDF
   ISN 0009              X=CHI
   ISN 0010              CHIPRB=1.0
   ISN 0011              IF(X.LE.0..OR.F.LT.1) RETURN
                C        LABEL *WRONG* IS OMITTED IN FAVOR OF RETURNING CHIPRB=1.0
   ISN 0013              A=0.5*X
   ISN 0014              BIGX=A.GT.10.
                C        BIGX SHOULD BE TRUE WHEN X IS SO BIG THAT EXP(-A) IS NOT ACCURATE.
   ISN 0015              EVEN=(2*(F/2)-F).EQ.0
   ISN 0016              IF(EVEN.OR.(F.GT.2.AND..NOT.BIGX)) Y=EXP(-A)
   ISN 0018              IF(EVEN) S=Y
   ISN 0020              IF(.NOT.EVEN) S=2.0*NORMAL(-SQRT(X))
   ISN 0022              CHIPRB=S
   ISN 0023              IF(F.LE.2) RETURN
   ISN 0025              X=0.5*(F-1.0)
   ISN 0026              IF(EVEN) Z=1.0
   ISN 0028              IF(.NOT.EVEN) Z=0.5
```

```
ISN 0030          IF(.NOT.BIGX) GO TO 2
ISN 0032          IF(EVEN) E=0.
ISN 0034          IF(.NOT.EVEN) E=0.5723649
ISN 0036          C=ALOG(A)
ISN 0037     1    E=ALOG(Z)+E
ISN 0038          S=EXP(C*Z-A-E)+S
ISN 0039          Z=Z+1.0
ISN 0040          IF(Z.LE.X) GO TO 1
ISN 0042          CHIPRB=S
ISN 0043          RETURN
ISN 0044     2    IF(EVEN) E=1.0
ISN 0046          IF(.NOT.EVEN) E=0.5641896/SQRT(A)
ISN 0048          C=0.
ISN 0049     3    E=E*A/Z
ISN 0050          C=C+E
ISN 0051          Z=Z+1.0
ISN 0052          IF(Z.LE.X) GO TO 3
ISN 0054          CHIPRB=C*Y+S
ISN 0055          RETURN
ISN 0056          END

****** END OF COMPILATION ******
LEVEL 16 ( 1 JULY 68)                     OS/360  FORTRAN H                          DATE  69.172/05.05.15

        COMPILER OPTIONS - NAME=  MAIN,OPT=02,LINECNT=59,SOURCE,EBCDIC,NOLIST,NODECK,LOAD,NOMAP,NOEDIT,ID,NOXREF
ISN 0002          SUBROUTINE SIGDS(AL,NV,NF,NG,NT)
            C   PURPOSE       TEST SIGNIFICANCE OF ROOTS  HIS THE SUM OF THE ROOT
            C                 AND SMALLER ROOT IS EQUAL TO 0
            C                 SEE RAO  1965 P 474
            C       AL        INPUT VECTOR OF ROOTS
            C       NV        NO OF ORIGINAL VARIABLES
            C       NF        NO OF ROOTS TO BE TESTED, SHOULD BE LE NG-1
            C       NG        NO OF GROUPS
            C       NT        NO OF TOTAL OBSERVATIONS
ISN 0003          DIMENSION AL(1)
ISN 0004      100 FORMAT(1H0,I3,'TH ROOT=',F9.4,5X,'DF=',I5,5X,'CHI=',F10.4,5X,
                 1'PROBABILITY=',F9.6,5X,'PERCENT=',F6.3)
ISN 0005      101 FORMAT(/,/,1H0,'SIGNIFICANCE TEST OF ROOTS SEE RAO 1965 P 474')
ISN 0006          WRITE(6,101)
ISN 0007          TRACE=0.0
ISN 0008          DO 1 I=1,NV
ISN 0009        1 TRACE=TRACE+AL(I)
ISN 0010          DO 3 K=1,NF
ISN 0011          CHI=0.0
ISN 0012          DO 2 J=K,NV
ISN 0013        2 CHI=CHI+ALOG(1.0/(1.0+AL(J)))
ISN 0014          CHI=-(NT-NG-(NV+NG)/2.0)*CHI
ISN 0015          NDF=(NV-K+1)*(NG-K)
ISN 0016          PRO=CHIPRB(CHI,NDF)
ISN 0017          PER=(100*AL(K))/TRACE
ISN 0018        3 WRITE(6,100) K,AL(K),NDF,CHI,PRO,PER
ISN 0019          RETURN
ISN 0020          END

****** END OF COMPILATION ******
LEVEL 16 ( 1 JULY 68)                     OS/360  FORTRAN H                          DATE  69.172/05.05.20

        COMPILER OPTIONS - NAME=  MAIN,OPT=02,LINECNT=59,SOURCE,EBCDIC,NOLIST,NODECK,LOAD,NOMAP,NOEDIT,ID,NOXREF
ISN 0002          SUBROUTINE DATRAN(N,X)
ISN 0003          DIMENSION X(1)
ISN 0004          RETURN
ISN 0005          END

****** END OF COMPILATION ******

U OF A   JOB STATISTICS --    449 CARDS READ --     505 LINES PRINTED --     0 CARDS PUNCHED --   1.03 MINUTES EXECUTION TIME
```

```
LEVEL 16 ( 1 JULY 68)                    OS/360  FORTRAN H                                DATE  69.171/12.59.20

        COMPILER OPTIONS - NAME=  MAIN,OPT=02,LINECNT=59,SOURCE,EBCDIC,NOLIST,NODECK,LOAD,NOMAP,NOEDIT,ID,NOXREF
              C        DIVISION OF EDUCATIONAL RESEARCH SERVICES
              C        UNIVERSITY OF ALBERTA
              C   MULV11
              C   PURPOSE:      CLASSIFY SUBJECTS BASED ON PROBABILITY CALCULATED BY
              C                 DISCRIMINANT ANALYSIS
              C
              C
              C   CARD INPUT
              C        1)          TITELE(20A4)
              C        2)          PARAMETERS(16I5) NP,NG,NV,NF,IPROB,NID
              C        3)          FORMAT FOR INPUT MATRICES(20A4)
              C        4)          C-NGXNV CENTROID MATRIX IN ORIGINAL SPACE
              C        5)          D-NVXNV DISPERSION MATRIX IN ORIGINAL SPACE
              C        6)          A-NVXNF TRANSFORMATION MATRIX
              C                    (4,5,6 ARE OUT PUT FROM MULV10 PROGRAM)
              C        7)          IF IPROB=1, NO OF OBSERVATIONS IN EACH GROUP USED FOR
              C                    POPULATION STUDY (16I5)
              C                    IF IPROB=2 INPUT A PRIORI PROBABILITY FOR EACH GROUP
              C                    (16F5.5)
              C        8)          FORMAT FOR THE DATA(20A4)
              C        9)          INPUT SAMPLE DATA
              C       10)          A BLANK CARD
              C   PARAMETERS
              C        NP          NO OF PERSON TO BE CLASSIFIED
              C        NG          NO OF GROUPS
              C        NV          NO OF VARIABLES IN ORIGINAL SPACE
              C        NF          NO OF VARIABLES(FACTORS) AFTER TRANSFORMATION
              C        IPROB       INDICATES A PRIORI PROBABILITY TYPE FOR APPLYING
              C                    BAYES RULE TO CALCULATE A POSTERIOR PROBABILITY WHEN
              C                    THE TEST SCORES ARE GIVEN FOR EACH INDIVIDUAL TO BE
              C                    CLASSIFIED
              C                    1-PROPORTIONAL TO NO OF OBSERVATION IN EACH GROUP
              C                    USED TO OBTAIN  INPUT MATRICES
              C                    2-TO BE SPECIFIED BY CARD INPUT(10F8.6)
              C                    IF NOT 1 OR 2 ASSUME EQUAL PROPABILITY FOR ALL GROUPS
              C        NID         IF NID=1 INDICATE THERE ARE ID NUMBERS FOR INPUT DATA
              C                    CARD(6A1),OTHERWISE ASSUME NO ID NUMBER
              C   SUBPROGRAMS      CLASSI,MXOUT,CHIPRB
              C                SSP ARRAY,CCPY,GMPRD,GMTRA,GTPRD
              C   PROGRAMMER       K. BAY
   ISN 0002          DIMENSION TITLE(20),FMT(20),A(50,25),D(50,50),C(50,50),CC(25,50),
                    1DD(25,25),S(50,50),P(50),PP(50),NCELL(50),LL(50),MM(50),SS(50,50),
                    2SSS(50,50),DATA(50)
   ISN 0003      100 FORMAT(20A4)
   ISN 0004      101 FORMAT(1H1,10X,20A4)
   ISN 0005      102 FORMAT(16I5)
   ISN 0006      103 FORMAT(1H0,'FORMAT FOR TRANSFORMATION,DISPEPSION, AND CENTROID MAT
                    1RIX')
   ISN 0007      104 FORMAT(1X,20A4)
   ISN 0008      105 FORMAT(1X,'FORMAT FOR THE DATA')
   ISN 0009      106 FORMAT(1H0, 'NO OF SUBJECTS',19X,I5,/,1X,'NO OF GROUPS',21X,I5,/,
                    11X,'NO OF VARIABLES',18X,I5,/,1X,'NO OF VARIABLES IN REDUCED SPACE
                    2',1X,I5)
   ISN 0010      107 FORMAT(1H0,'NO OF OBSERVATION IN EACH CELL USED FOR ORIGINAL DATA'
                    1)
   ISN 0011      108 FORMAT(1X,16I5)
   ISN 0012      109 FORMAT(16F5.5)
   ISN 0013      110 FORMAT(1H0,'NO OF OBSERVATION IN EACH GROUP IN POPULATION STUDY')
   ISN 0014        1 READ(5,100,END=9999) TITLE
   ISN 0015          IF(TITLE(1).EQ.TITLE(2)) GO TO 9999
   ISN 0017          WRITE(6,101) TITLE
   ISN 0018          READ(5,102) NP,NG,NV,NF,IPROB,NID
   ISN 0019          WRITE(6,106)NP,NG,NV,NF
   ISN 0020          READ(5,100) FMT
   ISN 0021          WRITE(6,103)
   ISN 0022          WRITE(6,104) FMT
   ISN 0023          DO 2 I=1,NG
   ISN 0024        2 READ(5,FMT) (C(I,J),J=1,NV)
   ISN 0025          CALL ARRAY(2,NG,NV,50,50,C,C)
   ISN 0026          CALL MXOUT(C,NG,NV,0,60,132,1,40,40H INPUT CENTROID MATRIX GROUPS
                    1BY VARS    )
   ISN 0027          DO 3 I=1,NV
   ISN 0028        3 READ(5,FMT) (D(I,J),J=1,NV)
   ISN 0029          CALL ARRAY(2,NV,NV,50,50,D,D)
   ISN 0030          CALL MXOUT(D,NV,NV,0,60,132,1,24,24H INPUT DISPERSION MATRIX)
   ISN 0031          DO 4 I=1,NV
   ISN 0032        4 READ(5,FMT) (A(I,J),J=1,NF)
   ISN 0033          CALL ARRAY(2,NV,NF,50,25,A,A)
   ISN 0034          CALL MXOUT(A,NV,NF,0,60,132,1,28,28H INPUT TRANSFORMATION MATRIX)
   ISN 0035          CALL GMPRD(C,A,S,NG,NV,NF)
   ISN 0036          CALL GMTRA(S,CC,NG,NF)
   ISN 0037          CALL MXOUT(CC ,NF,NG,0,60,132,1,52,52H CENTROID MATRIX IN REDUCED
                    1SPACE FACTOR BY GROUPS  )
   ISN 0038          CALL GTPRD(A,D,S,NV,NF,NV)
   ISN 0039          CALL GMPRD(S,A,DD,NF,NV,NF)
   ISN 0040          CALL MXOUT(DD,NF,NF,0,60,132,1,36,36H DISPERSION MATRIX IN REDUCED
                    1 SPACE   )
   ISN 0041          IF (IPROB.NE.1) GO TO 11
   ISN 0043          READ(5,102) (NCELL(I),I=1,NG)
   ISN 0044          WRITE(6,110)
   ISN 0045          WRITE(6,108) (NCELL(I),I=1,NG)
   ISN 0046          FCELL=0.0
   ISN 0047          DO 5 I=1,NG
   ISN 0048        5 FCELL=FCELL+NCELL(I)
   ISN 0049          DO 6 I=1,NG
   ISN 0050        6 P(I)=NCELL(I)/FCELL
   ISN 0051          GO TO 10
   ISN 0052       11 CONTINUE
   ISN 0053        7 IF(IPROB.NE.2) GO TO 9
   ISN 0055          READ(5,109) (P(I),I=1,NG)
   ISN 0056          GO TO 10
   ISN 0057        9 PPP=1.0/NG
   ISN 0058          DO 8 I=1,NG
   ISN 0059        8 P(I)=PPP
   ISN 0060       10 CALL MXOUT(P,1,NG,0,60,132,1,24,24H PRIORI PROBABILITY USED)
   ISN 0061          READ(5,100) FMT
```

```
ISN 0062          WRITE(6,105)
ISN 0063          WRITE(6,104) FMT
ISN 0064          CALL MINV(OO,NF,DET,LL,MM)
ISN 0065          CALL MXOUT(OO,NF,NF,0,60,132,1,32,32H INVERSE OF DISPERSION MATRIX
                 1    )
ISN 0066          WRITE(6,101)
ISN 0067          CALL CLASSI(NP,NG,NV,NF,NCELL,P,A,CC,FMT,OO,PP,S,SS,SSS,DATA,NIO)
ISN 0068          GO TO 1
ISN 0069     9999 STOP
ISN 0070          END

****** END OF COMPILATION ******
LEVEL 16 ( 1 JULY 68)                    OS/360  FORTRAN H                    DATE  69.171/12.59.39

       COMPILER OPTIONS - NAME=  MAIN,OPT=02,LINECNT=59,SOURCE,EBCDIC,NOLIST,NODECK,LOAD,NOMAP,NOEDIT,ID,NOXREF
ISN 0002          SUBROUTINE CLASSI(NP,NG,NV,NF,NCELL,P,FL,CENT,FMT,O,PP,CHI,XX,GM,
                 1OO,NIO)
            C   PURPOSE :
            C        CLASSIFY SUBJECTS BASED ON NV OBSERVATION AND POPULATION
            C        CENTROID, DISPERSION IN REDUCED SPACE AND TRANSFORMATION
            C        MATRICES
            C        NP         NO OF PERSONS TO BE CLLASIFIED
            C        NG         NO OF GROUPS
            C        NV         NO OF INPUT VARIABLES .
            C        NF         NO OF VARIABLES IN REDUCED SPACE, USUALLY NG-1
            C        P          INPUT VECTOR INDICATING A PRIORI PROBABILITY FOR EACH
            C                   GROUP
            C        FL         INPUT TRANSFORMATION MATRIX OF SIZE NVXNF
            C        CENT       INPUT NFXNG CENTROID MATRIX
            C        FMT        FORMAT FOR THE DATA
            C        O          INVERSE OF DISPERSION MATRIX IN REDUCED SPACE
            C        LL,MM      WORKING VECTORS
            C        OO,CHI     WORKING VECTORS OF SIZE NV AND NG
            C        NID        IF EQ TO 1 ID NO IS INPUT FOR EACH SUBJECT,
            C                   OTHERWISE ASSUME NO ID INPUT
ISN 0003          DIMENSION NCELL(1),P(1),PP(1),FL(1),CENT(1),FMT(1),O(1),OO(1),
                 1XX(1),GM(1),CHI(1)
ISN 0004          DIMENSION ID(6)
ISN 0005          DATA ID/'*','*','*','*','*','*'/
ISN 0006      100 FORMAT(1X,8X,'CHISQ',8X,10(1X,F8.4),/, 1X,21X,10(1X,F8.4))
ISN 0007      101 FORMAT(1X,8X,'POSTERIOR P',2X,10(1X,F8.4),/,2(1X,21X,10(1X,F8.4),
                 1/,))
ISN 0008      102 FORMAT(1H0,6A1,2X,'SCORE',8X,10(1X,F8.4))
ISN 0009      103 FORMAT(1X,8X,'CONDITION P',2X,10(1X,F8.4),/,2(1X,21X,10(1X,F8.4),
                 1/,))
ISN 0010      104 FORMAT(1X,'.....CLASSIFICATION.....',5X,'CONDITIONAL:',2X,I2,5X,
                 1'BAYSIAN RULE:',2X,I2,1X,'...............................')
ISN 0011          DO 11 I=1,NP
ISN 0012          IF (NIO.EQ.1) GO TO 1
ISN 0014          READ(5,FMT)     (OO(J),J=1,NV)
ISN 0015          GO TO 2
ISN 0016        1 READ(5,FMT) (ID(K),K=1,6),(OO(J),J=1,NV)
ISN 0017        2 CALL GMPRO(OO,FL,XX,1,NV,NF)
ISN 0018          WRITE(6,102) (ID(K),K=1,6),(XX(J),J=1,NF)
ISN 0019          DO 5 M=1,NG
ISN 0020          CALL CCPY(CENT,M,GM,NF,NG,0)
ISN 0021          DO 3 L=1,NF
ISN 0022        3 GM(L)=XX(L)-GM(L)
ISN 0023          CALL GMPRO(O,GM,OO,NF,NF,1)
ISN 0024          X2=0.0
ISN 0025          DO 4 J=1,NF
ISN 0026        4 X2=X2+GM(J)*OD(J)
ISN 0027        5 CHI(M)=X2
ISN 0028          SUM=0.0
ISN 0029          DO 6 J=1,NG
ISN 0030          PP(J)=P(J)*EXP(-CHI(J)/2.0)
ISN 0031        6 SUM=SUM+PP(J)
ISN 0032          DO 7 J=1,NG
ISN 0033        7 PP(J)=PP(J)/SUM
ISN 0034          WRITE(6,100)     (CHI(J),J=1,NG)
ISN 0035          WRITE(6,101) (PP(J),J=1,NG)
ISN 0036          NCB=1
ISN 0037          DO 8 J=2,NG
ISN 0038        8 IF(PP(J).GT.PP(NCB)) NCB=J
ISN 0040          DO 9 J=1,NG
ISN 0041        9 PP(J)=CHIPRB(CHI(J),NF)
ISN 0042          WRITE(6,103) (PP(J),J=1,NG)
ISN 0043          NCA=1
ISN 0044          DO 10 J=2,NG
ISN 0045       10 IF(PP(J).GT.PP(NCA)) NCA=J
ISN 0047          WRITE(6,104) NCA,NCB
ISN 0048       11 CONTINUE
ISN 0049          RETURN
ISN 0050          END

****** END OF COMPILATION ******
LEVEL 16 ( 1 JULY 68)                    OS/360  FORTRAN H                    DATE  69.171/12.59.55

       COMPILER OPTIONS - NAME=  MAIN,OPT=02,LINECNT=59,SOURCE,EBCDIC,NOLIST,NODECK,LOAD,NOMAP,NOEDIT,ID,NOXREF
ISN 0002          SUBROUTINE MXOUT(A,N,M,MS,LINS,IPOS,ISP,NUMHOL,TITLE)
            C
            C        A        :NAME OF OUTPUT MATRIX
            C        N        :NUMBER OF ROWS IN A
            C        M        :NUMBER OF COLUMNS IN A
            C        MS       :STORAGE MODE OF A
            C                    0 GENERAL
            C                    1 SYMETRIC
            C                    2 DIAGONAL
            C        LINS     :NUMBER OF PRINT LINES ON THE PAGE(USUALLY 60)
            C        IPOS     :NUMBER OF PRINT POSITIONS ACROSS THE PAGE(USUALLY 132)
            C        ISP      :LINE SPACING CODE, 1 FOR SINGLE SPACE 2 FOR DOUBLE
            C                  SPACE
ISN 0003          DIMENSION A(1),B(8),TITLE(20)
ISN 0004        1 FORMAT (1H0,/,1H0,20A4)
ISN 0005        2 FORMAT(12X,8HCOLUMN  ,7(3X,I3,10X))
ISN 0006        3 FORMAT(1H )
ISN 0007        4 FORMAT(1H ,7X,4HROW ,I3,7(E16.6))
ISN 0008        5 FORMAT(1H0,7X,4HROW ,I3,7(E16.6))
```

```
ISN 0009          NN=(NUMHOL+3)/4
ISN 0010          WRITE(6,1) (TITLE(J),J=1,NN)
ISN 0011          J=1
ISN 0012          NEND=IPOS/16-1
ISN 0013          LENO=(LINS/ISP)-2
ISN 0014       10 LSTRT=1
ISN 0015       20 CONTINUE
ISN 0016          JNT=J+NENO-1
ISN 0017       31 IF(JNT-M)33,33,32
ISN 0018       32 JNT=M
ISN 0019       33 CONTINUE
ISN 0020          WRITE(6,2) (JCUR,JCUR=J,JNT)
ISN 0021          IF(ISP-1) 35,35,40
ISN 0022       35 WRITE(6,3)
ISN 0023       40 LTENO=LSTRT+LENO-1
ISN 0024          DO 80 L=LSTRT,LTEND
ISN 0025          DO 55 K=1,NENO
ISN 0026          KK=K
ISN 0027          JT=J+K-1
ISN 0028          CALL LOC(L,JT,IJNT,N,M,MS)
ISN 0029          B(K)=0.0
ISN 0030          IF(IJNT) 50,50,45
ISN 0031       45 B(K)=A(IJNT)
ISN 0032       50 CONTINUE
ISN 0033          IF(JT-M) 55,60,60
ISN 0034       55 CONTINUE
ISN 0035       60 IF(ISP-1) 65,65,70
ISN 0036       65 WRITE(6,4) L,(B(JW),JW=1,KK)
ISN 0037          GO TO 75
ISN 0038       70 WRITE(6,5) L,(B(JW),JW=1,KK)
ISN 0039       75 IF(N-L) 85,85,80
ISN 0040       80 CONTINUE
ISN 0041          LSTRT=LSTRT+LENO
ISN 0042          GO TO 20
ISN 0043       85 IF(JT-M) 90,95,95
ISN 0044       90 J=JT+1
ISN 0045          GO TO 10
ISN 0046       95 RETURN
ISN 0047          END

****** END OF COMPILATION ******
LEVEL 16 ( 1 JULY 68)                    OS/360  FORTRAN H                    DATE  69.171/13.00.05

       COMPILER OPTIONS - NAME=  MAIN,OPT=02,LINECNT=59,SOURCE,EBCDIC,NOLIST,NODECK,LOAD,NOMAP,NOEDIT,ID,NOXREF
ISN 0002          FUNCTION CHIPRB(CHI,NDF)
             C    COMPUTES PROBABILITY OF CHISQUARE CHI WITH NDF DEGREES OF FREEDOM.
ISN 0003          EXTERNAL SQRT,ERF
ISN 0004          REAL NORMAL
ISN 0005          INTEGER F
ISN 0006          LOGICAL BIGX,EVEN
ISN 0007          NORMAL(X)=0.5*(1.0+ERF(0.7071068*(-SQRT(X))))
             C    SEE KENNEY AND KEEPING(1951) VOL.2, P.43, WHERE ERF IS CALLED G.
ISN 0008          F=NDF
ISN 0009          X=CHI
ISN 0010          CHIPRB=1.0
ISN 0011          IF(X.LE.0..OR.F.LT.1) RETURN
             C    LABEL *WRONG* IS OMITTED IN FAVOR OF RETURNING CHIPRB=1.0
ISN 0013          A=0.5*X
ISN 0014          BIGX=A.GT.10.
             C    BIGX SHOULD BE TRUE WHEN X IS SO BIG THAT EXP(-A) IS NOT ACCURATE.
ISN 0015          EVEN=(2*(F/2)-F).EQ.0
ISN 0016          IF(EVEN.OR.(F.GT.2.AND..NOT.BIGX)) Y=EXP(-A)
ISN 0018          IF(EVEN) S=Y
ISN 0020          IF(.NOT.EVEN) S=2.0*NORMAL(X)
ISN 0022          CHIPRB=S
ISN 0023          IF(F.LE.2) RETURN
ISN 0025          X=0.5*(F-1.0)
ISN 0026          IF(EVEN) Z=1.0
ISN 0028          IF(.NOT.EVEN) Z=0.5
ISN 0030          IF(.NOT.BIGX) GO TO 2
ISN 0032          IF(EVEN) E=0.
ISN 0034          IF(.NOT.EVEN) E=0.5723649
ISN 0036          C=ALOG(A)
ISN 0037       1  E=ALOG(Z)+E
ISN 0038          S=EXP(C*Z-A-E)+S
ISN 0039          Z=Z+1.0
ISN 0040          IF(Z.LE.X) GO TO 1
ISN 0042          CHIPRB=S
ISN 0043          RETURN
ISN 0044       2  IF(EVEN) E=1.0
ISN 0046          IF(.NOT.EVEN) E=0.5641896/SQRT(A)
ISN 0048          C=0.
ISN 0049       3  E=E*A/Z
ISN 0050          C=C+E
ISN 0051          Z=Z+1.0
ISN 0052          IF(Z.LE.X) GO TO 3
ISN 0054          CHIPRB=C*Y+S
ISN 0055          RETURN
ISN 0056          END

****** END OF COMPILATION ******

U OF A   JOB STATISTICS --   291 CARDS READ --   330 LINES PRINTED --   0 CARDS PUNCHED --   0.53 MINUTES EXECUTION TIME
```

```
LEVEL 16 ( 1 JULY 68)                    OS/360  FORTRAN H                    DATE  69.171/12.59.23

       COMPILER OPTIONS - NAME=  MAIN,OPT=02,LINECNT=59,SOURCE,EBCDIC,NOLIST,NODECK,LOAD,NOMAP,NOEDIT,ID,NOXREF
            C        DIVISION OF EDUCATIONAL RESEARCH SERVICES
            C        UNIVERSITY OF ALBERTA
            C   MULV12
            C   PURPOSE:     CALCULATE DISCRIMINANT SCORES AND CLASSIFY SUBJECTS
            C                BASED ON CONDITIONAL PROBABILITY AND BAYES RULE
            C                USING EACH GROUP DISPERSION MATRIX
            C
            C   CARD INPUT
            C        1)          TITLE(20A4)
            C        2)          PARAMETERS(16I5) NP,NG,NV,NF,IPROB,NID
            C            NP      NO OF PERSON TO BE CLASSIFIED
            C            NG      NO OF GROUPS
            C            NV      NO OF VARIABLES IN ORIGINAL SPACE
            C            NF      NO OF VARIABLES(FACTORS) AFTER TRANSFORMATION
            C            IPROB   INDICATES A PRIORI PROBABILITY TYPE FOR APPLYING
            C                    BAYES RULE TO CALCULATE A POSTERIOR PROBABILITY WHEN
            C                    THE TEST SCORES ARE GIVEN FOR EACH INDIVIDUAL TO BE
            C                    CLASSIFIED
            C                    1-PROPORTIONAL TO NO OF OBSERVATION IN EACH GROUP
            C                    USED TO OBTAIN  INPUT MATRICES
            C                    2-TO BE SPECIFIED BY CARD INPUT
            C                    IF NOT 1 OR 2 ASSUME EQUAL PROPABILITY FOR ALL GROUPS
            C            NID     IF NID=1 INDICATE THERE ARE ID NUMBERS FOR INPUT DATA
            C                    CARD(6A1),OTHERWISE ASSUME NO ID NUMBER
            C        3)          FORMAT FOR INPUT MATRICES(20A4)
            C        4)          C-NGXNV CENTROID MATRIX IN ORIGINAL SPACE
            C        5)          D-NVXNV DISPERSION MATRIX IN ORIGINAL SPACE
            C        6)          A-NVXNF TRANSFORMATION MATRIX
            C        7)          NG GROUP DISPERSION MATRIX OF SIZE NVXNV
            C                    (4,5,7 ARE OUT PUT FROM MULV10 PROGRAM)
            C        8)          IF IPROB=1, NO OF OBSERVATIONS IN EACH GROUP USED FOR
            C                    POPULATION STUDY (16I5)
            C                    IF IPROB=2 INPUT A PRIORI PROBABILITY FOR EACH GROUP
            C                    (16F5.5)
            C        9)          FORMAT FOR THE DATA(20A4)
            C       10)          INPUT SAMPLE DATA
            C       11)          A BLANK CARD
            C   SUBPROGRAMS      CLASSI,MXOUT,CHIPRB
            C                SSP ARRAY,CCPY,GMPRD,GMTRA,GTPRD,MINV,MPRD,MSTR,LOC
            C   PROGRAMMER       K. BAY
            C   REMARKS          1) CURRENTLY DIMENSIONED TO ACCOMODATE UP TO 50
            C                       VARIABLES,25 GROUPS,AND 24 DISCRIMINANT SCORES
  ISN 0002       DIMENSION TITLE(20),FMT(20),A(50,25),D(50,50),C(50,50),CC(25,50),
                1DD(25,25),S(2500 ),P(50),PP(50),NCELL(50),LL(50),MM(50),SS(50,50),
                2SSS(50,50),DATA(50),DG(300,25),DP(50)
  ISN 0003   100 FORMAT(20A4)
  ISN 0004   101 FORMAT(1H1,10X,20A4)
  ISN 0005   102 FORMAT(16I5)
  ISN 0006   103 FORMAT(1H0,'FORMAT FOR TRANSFORMATION,DISPERSION, AND CENTROID MAT
                1RIX')
  ISN 0007   104 FORMAT(1X,20A4)
  ISN 0008   105 FORMAT(1X,'FORMAT FOR THE DATA')
  ISN 0009   106 FORMAT(1H0, 'NO OF SUBJECTS',19X,I5,/,1X,'NO OF GROUPS',21X,I5,/,
                11X,'NO OF VARIABLES',18X,I5,/,1X,'NO OF VARIABLES IN REDUCED SPACE
                2',1X,I5)
  ISN 0010   107 FORMAT(1H0,'NO OF OBSERVATION IN EACH CELL USED FOR ORIGINAL DATA'
                1)
  ISN 0011   108 FORMAT(1X,16I5)
  ISN 0012   109 FORMAT(16F5.5)
  ISN 0013   110 FORMAT(1H0,'NO OF OBSERVATION IN EACH GROUP IN POPULATION STUDY')
  ISN 0014   111 FORMAT(/,/,1H0,'GROUP DISPERSIONS')
  ISN 0015   112 FORMAT(1H0,'GROUP :',I3)
  ISN 0016     1 READ(5,100,END=9999) TITLE
  ISN 0017       IF(TITLE(1).EQ.TITLE(2)) GO TO 9999
  ISN 0019       WRITE(6,101) TITLE
  ISN 0020       READ(5,102) NP,NG,NV,NF,IPROB,NID
  ISN 0021       WRITE(6,106)NP,NG,NV,NF
  ISN 0022       READ(5,100) FMT
  ISN 0023       WRITE(6,103)
  ISN 0024       WRITE(6,104) FMT
  ISN 0025       DO 2 I=1,NG
  ISN 0026     2 READ(5,FMT) (C(I,J),J=1,NV)
  ISN 0027       CALL ARRAY(2,NG,NV,50,50,C,C)
  ISN 0028       CALL MXOUT(C,NG,NV,0,60,132,1,40,40H INPUT CENTROID MATRIX GROUPS
                1BY VARS    )
  ISN 0029       DO 3 I=1,NV
  ISN 0030     3 READ(5,FMT) (D(I,J),J=1,NV)
  ISN 0031       CALL ARRAY(2,NV,NV,50,50,D,D)
  ISN 0032       CALL MXOUT(D,NV,NV,0,60,132,1,24,24H INPUT DISPERSION MATRIX)
  ISN 0033       DO 4 I=1,NV
  ISN 0034     4 READ(5,FMT) (A(I,J),J=1,NF)
  ISN 0035       CALL ARRAY(2,NV,NF,50,25,A,A)
  ISN 0036       CALL MXOUT(A,NV,NF,0,60,132,1,28,28H INPUT TRANSFORMATION MATRIX)
  ISN 0037       CALL GMPRD(C,A,S,NG,NV,NF)
  ISN 0038       CALL GMTRA(S,CC,NG,NF)
  ISN 0039       CALL MXOUT(CC ,NF,NG,0,60,132,1,52,52H CENTROID MATRIX IN REDUCED
                1SPACE FACTOR BY GROUPS   )
  ISN 0040       CALL GTPRD(A,D,S,NV,NF,NV)
  ISN 0041       CALL GMPRD(S,A,DD,NF,NV,NF)
  ISN 0042       CALL MXOUT(DD,NF,NF,0,60,132,1,36,36H DISPERSION MATRIX IN REDUCED
                1 SPACE   )
  ISN 0043       CALL MINV(DD,NF,DET,LL,MM)
  ISN 0044       CALL MXOUT(DD,NF,NF,0,60,132,1,32,32H INVERSE OF DISPERSION MATRIX
                1    )
  ISN 0045       NFH=((NF+1)*NF)/2
  ISN 0046       WRITE(6,111)
  ISN 0047       DO 14 IG=1,NG
  ISN 0048       WRITE(6,112) IG
  ISN 0049       DO 12 I=1,NV
  ISN 0050    12 READ(5,FMT) (D(I,J),J=1,NV)
  ISN 0051       CALL ARRAY(2,NV,NV,50,50,D,D)
  ISN 0052       CALL MXOUT(D,NV,NV,0,60,132,1,20,20H IN ORIGINAL SPACE    )
  ISN 0053       CALL GTPRD(A,D,S,NV,NF,NV)
  ISN 0054       CALL GMPRD(S,A,D,NF,NV,NF)
  ISN 0055       CALL MXOUT(D,NF,NF,0,60,132,1,20,20H IN REDUCED SPACE    )
```

```
ISN 0056            CALL MINV(D,NF,OET,LL,MM)
ISN 0D57          . DPIIG)=DET
ISN 0058            CALL MXOUTID,NF,NF,0,6D,132,1,24,24H INVERSE OF OISPERSION  )
ISN 0059            CALL MSTR(O,S,NF,0,1)
ISN 0060            OO 13 J=1,NFH
ISN 0061         13 DG(J,IG)=S(J)
ISN 0062         14 CONTINUF
ISN 0063            WRITE(6,101)
ISN 0064            CALL ARRAY(2,NFH,NG,30D,25,OG,DG)
ISN 0065            CALL MXOUT(OP,1,NG,0,60,132,1,20,20H DETERMINANTS        )
ISN 0066            IF (IPROB.NE.1) GO TO 1I
ISN 0068            READ(5,102) (NCELL(I),I=1,NG)
ISN 0069            WRITE(6,11D)
ISN 0D70            WRITE(6,108) (NCELL(I),I=1,NG)
ISN 0071            FCFLL=0.0
ISN 0D72            OO 5 I=1,NG
ISN 0073          5 FCELL=FCELL+NCELL(I)
ISN 0074            DO 6 I=1,NG
ISN 0075          6 P(I)=NCELL(I)/FCELL
ISN 0076            GO TO 10
ISN 0077         11 CONTINUE
ISN 0D78          7 (F(IPRO8.NE.2) GO TO 9
ISN 0080            READ(5,109) IP(1),I=1,NG)
ISN 0D81            GO TO 10
ISN 0082          9 PPP=1.0/NG
ISN D083            DO 8 I=1,NG
ISN 0084          8 P(I)=PPP
ISN 0D85         10 CALL MXOUT(P,1,NG,0,60,132,1,24,24H PRIORI PROBABILITY USEO)
ISN 0086            READ(5,100) FMT
ISN 0D87            WR(TEI6,105)
ISN 0D88            WRITF(6,104) FMT
ISN 0089            WRITE(6,101)
ISN 0090            CALL CLASS1(NP,NG,NV,NF,NCELL,P,A,CC,FMT,DD,PP,S,SS,SSS,DATA,NID,
                   1OG,DP)
ISN D091            GO TO 1
ISN 0D92       9999 STOP
ISN 0093            END

****** END OF COMPILATION ******
LEVEL 16 I 1 JULY 68)                       OS/360  FORTRAN H                          OATE  69.171/12.59.37

         COMPILER OPTIONS - NAME=  MAIN,OPT=02,LINFCNT=59,SOURCE,E8COIC,NOLIST,NOOECK,LOAD,NOMAP,NOEOIT,IO,NOXREF
ISN 0002            SUBROUTINE CLASSIINP,NG,NV,NF,NCELL,P,FL,CENT,FMT,D,PP,CHI,XX,GM,
                   1OD,NIO,OG,OP)
               C   PURPOSE :
               C        CLASSIFY SU8JECTS 8ASED ON NV O8SERVATION ANO POPULATION
               C        CENTROID, OISPERSION IN REOUCED SPACE AND TRANSFORMATION
               C        MATRICES
               C        NP          NO OF PERSONS TO 8E CLLASIFIED
               C        NG          NO OF GROUPS
               C        NV          NU OF INPUT VARIABLES
               C        NF          NO OF VARIABLES IN REDUCED SPACE, USUALLY NG-1
               C        P           INPUT VECTOP INOICATING A PRIORI PROBABILITY FOR EACH
               C                    GROUP
               C        FL          INPUT TRANSFORMATION MATRIX OF SIZF NVXNF
               C        CENT        INPUT NFXNG CENTROIO MATRIX
               C        FMT         FORMAT FOR THE DATA
               C        D           INVERSE OF OISPERSION MATRIX IN REDUCED SPACE
               C        LL,MM       WORKING VECTORS
               C        OO,CHI      WORKING VECTORS OF SIZF NV AND NG
               C        NIO         IF EQ TO 1 ID NO IS INPUT FOR FACH SUBJECT,
               C                    OTHERWISE ASSUME NO IO INPUT
ISN 0003            DIMENSION NCELL(1),P(1),PPI1),FL(1),CENT(1),FMTI1),O(1),OO(1),
                   1XX(1),GMI1),CHI(1),OGI1),DP(1)
ISN 0004            DIMENSION IO(6)
ISN 0005            DATA IO/'*','*','*','*','*','*'/
ISN 0006        1D0 FORMAT(1X,8X,'CHISQ',8X,10(1X,F8.4),/, 1X,21X,10(1X,F8.4))
ISN 0007        101 FORMAT(1X,8X,'POSTERIOR P',2X,10I1X,F8.4),/,2I1X,21X,10(1X,F8.4),
                   1/,))
ISN 0008        102 FORMATI1HO,6A1,2X,'SCORE',8X,10(1X,F8.4))
ISN 0009        103 FORMAT(1X,8X,'CONDITION P',2X,10(1X,F8.4),/,2(1X,21X,10I1X,F8.4),
                   1/,))
ISN 0010        104 FORMAT(1X,'.....CLASSIFICATION.....',5X,'CONOITIONAL:',2X,I2,5X,
                   1'BAYSIAN RULE:',2X,I2,1X,'..............................')
ISN 0011            NFH=(INF+1)*NF)/2
ISN 0012            DO 11 MM=1,NP
ISN 0013            IF (NIO.EQ.1) GO TO 1
ISN 0015            READ(5,FMT)    IDO(J),J=1,NV)
ISN 0016            GO TO 2
ISN D017          I READI5,FMT) IIDIK),K=1,6),(DDIJ),J=1,NV)
ISN 001P          2 CALL GMPRO(DO,FL,XX,1,NV,NF)
ISN 0019            WRITE(6,102) IIOIK),K=1,6),(XXIJ),J=1,NF)
ISN DD2D            DO 5 M=1,NG
ISN 0021            CALL CCPY(CENT,M,GM,NF,NG,0)
ISN 0022            CALL CCPY(OG,M,O,NFH,NG,0)
ISN D023            DO 3 L=1,NF
ISN 0024          3 GMIL)=XX(L)-GMIL)
ISN 0D25            CALL MPRO(D,GM,OD,NF,NF,1,0,1)
ISN 0026            X2=0.0
ISN 0027            OO 4 J=1,NF
ISN 0028          4 X2=X2+GM(J)*OO(J)
ISN 0029          5 CHI(M)=X2
ISN 0030            SUM=0.0
ISN 0031            OO 6 J=1,NG
ISN 0032            PPIJ)=IPIJ)*EXP(-CHI(J)/2.0))/ISQRT(OPIJ)))
ISN 0033          6 SUM=SUM+PP(J)
ISN 0034            DO 7 J=1,NG
ISN 0035          7 PP(J)=PP(J)/SUM
ISN 0036            WRITE(6,100)    (CHIIJ),J=1,NG)
ISN D037            WRITE(6,101) IPP(J),J=1,NG)
ISN 0038            NC8=1
ISN 0039            DO 8 J=2,NG
ISN 0040          8 IF(PPIJ).GT.PP(NCB)) NCB=J
ISN 0042            DO 9 J=1,NG
ISN D043          9 PPIJ)=CHIPRB(CHI(J),NF)
ISN 0044            WRITE(6,103) IPP(J),J=1,NG)
ISN 0D45            NCA=1
ISN D046            DO 10 J=2,NG
```

```
   ISN 0047         10 IF(PP(J).GT.PP(NCA)) NCA=J
   ISN 0049            WRITE(6,104) NCA,NCB
   ISN 0050         11 CONTINUE
   ISN 0051            RETURN
   ISN 0052            END

****** END OF COMPILATION ******
LEVEL 16 ( 1 JULY 68)                     OS/360  FORTRAN H                     DATE  69.171/12.59.48

      COMPILER OPTIONS - NAME=  MAIN,OPT=02,LINECNT=59,SOURCE,EBCDIC,NOLIST,NODECK,LOAD,NOMAP,NOEDIT,ID,NOXREF
   ISN 0002            SUBROUTINE MXOUT(A,N,M,MS,LINS,IPOS,ISP,NUMHOL,TITLE)
                C
                C         A       :NAME OF OUTPUT MATRIX
                C         N       :NUMBER OF ROWS IN A
                C         M       :NUMBER OF COLUMNS IN A
                C         MS      :STORAGE MODE OF A
                C                   0 GENERAL
                C                   1 SYMETRIC
                C                   2 DIAGONAL
                C         LINS    :NUMBER OF PRINT LINES ON THE PAGE(USUALLY 60)
                C         IPOS    :NUMBER OF PRINT POSITIONS ACROSS THE PAGE(USUALLY 132)
                C         ISP     :LINE SPACING CODE, 1 FOR SINGLE SPACE 2 FOR DOUBLE
                C                  SPACE
   ISN 0003            DIMENSION A(1),B(8),TITLE(20)
   ISN 0004          1  FORMAT (1H0,/,1H0,20A4)
   ISN 0005          2 FORMAT(12X,8HCOLUMN  ,7(3X,I3,10X))
   ISN 0006          3 FORMAT(1H )
   ISN 0007          4 FORMAT(1H ,7X,4HROW ,I3,7E16.6))
   ISN 0008          5 FORMAT(1H0,7X,4HROW ,I3,7E16.6))
   ISN 0009            NN=(NUMHOL+3)/4
   ISN 0010            WRITE(6,1) (TITLE(J),J=1,NN)
   ISN 0011            J=1
   ISN 0012            NEND=IPOS/16-1
   ISN 0013            LEND=(LINS/ISP)-2
   ISN 0014         10 LSTRT=1
   ISN 0015         20 CONTINUE
   ISN 0016            JNT=J+NEND-1
   ISN 0017         31 IF(JNT-M)33,33,32
   ISN 0018         32 JNT=M
   ISN 0019         33 CONTINUE
   ISN 0020            WRITE(6,2) (JCUR,JCUR=J,JNT)
   ISN 0021            IF(ISP-1) 35,35,40
   ISN 0022         35 WRITE(6,3)
   ISN 0023         40 LTEND=LSTRT+LEND-1
   ISN 0024            DO 80 L=LSTRT,LTEND
   ISN 0025            DO 55 K=1,NEND
   ISN 0026            KK=K
   ISN 0027            JT=J+K-1
   ISN 0028            CALL LOC(L,JT,IJNT,N,M,MS)
   ISN 0029            B(K)=0.0
   ISN 0030            IF(IJNT) 50,50,45
   ISN 0031         45 B(K)=A(IJNT)
   ISN 0032         50 CONTINUE
   ISN 0033            IF(JT-M) 55,60,60
   ISN 0034         55 CONTINUE
   ISN 0035         60 IF(ISP-1) 65,65,70
   ISN 0036         65 WRITE(6,4) L,(B(JW),JW=1,KK)
   ISN 0037            GO TO 75
   ISN 0038         70 WRITE(6,5) L,(B(JW),JW=1,KK)
   ISN 0039         75 IF(N-L) 85,85,80
   ISN 0040         80 CONTINUE
   ISN 0041            LSTRT=LSTRT+LEND
   ISN 0042            GO TO 20
   ISN 0043         85 IF(JT-M) 90,95,95
   ISN 0044         90 J=JT+1
   ISN 0045            GO TO 10
   ISN 0046         95 RETURN
   ISN 0047            END

****** END OF COMPILATION ******
LEVEL 16 ( 1 JULY 68)                     OS/360  FORTRAN H                     DATE  69.171/12.59.55

      COMPILER OPTIONS - NAME=  MAIN,OPT=02,LINECNT=59,SOURCE,EBCDIC,NOLIST,NODECK,LOAD,NOMAP,NOEDIT,ID,NOXREF
   ISN 0002            FUNCTION CHIPRB(CHI,NOF)
                C      COMPUTES PROBABILITY OF CHISQUARE CHI WITH NOF DEGREES OF FREEDOM.
   ISN 0003            EXTERNAL SQRT,ERF
   ISN 0004            REAL NORMAL
   ISN 0005            INTEGER F
   ISN 0006            LOGICAL BIGX,EVEN
   ISN 0007            NORMAL(X)=0.5*(1.0+ERF(0.7071068*(-SQRT(X))))
                C      SEE KENNEY AND KEEPING(1951) VOL.2, P.43, WHERE ERF IS CALLED G.
   ISN 0008            F=NOF
   ISN 0009            X=CHI
   ISN 0010            CHIPRB=1.0
   ISN 0011            IF(X.LE.0..OR.F.LT.1) RETURN
                C      LABEL *WRONG* IS OMITTED IN FAVOR OF RETURNING CHIPRB=1.0
   ISN 0013            A=0.5*X
   ISN 0014            BIGX=A.GT.10.
                C      BIGX SHOULD BE TRUE WHEN X IS SO BIG THAT EXP(-A) IS NOT ACCURATE.
   ISN 0015            EVEN=(2*(F/2)-F).EQ.0
   ISN 0016            IF(EVEN.OR.(F.GT.2.AND..NOT.BIGX)) Y=EXP(-A)
   ISN 0018            IF(EVEN) S=Y
   ISN 0020            IF(.NOT.EVEN) S=2.0*NORMAL(X)
   ISN 0022            CHIPRB=S
   ISN 0023            IF(F.LE.2) RETURN
   ISN 0025            X=0.5*(F-1.0)
   ISN 0026            IF(EVEN) Z=1.0
   ISN 0028            IF(.NOT.EVEN) Z=0.5
   ISN 0030            IF(.NOT.BIGX) GO TO 2
   ISN 0032            IF(EVEN) E=0.
   ISN 0034            IF(.NOT.EVEN) E=0.5723649
   ISN 0036            C=ALOG(A)
   ISN 0037          1 E=ALOG(Z)+E
   ISN 0038            S=EXP(C*Z-A-E)+S
   ISN 0039            Z=Z+1.0
   ISN 0040            IF(Z.LE.X) GO TO 1
   ISN 0042            CHIPRB=S
```

```
ISN 0043           RETURN
ISN 0044       2   IF(EVEN) E=1.0
ISN 0046           IF(.NOT.EVEN) E=0.5641896/SQRT(A)
ISN 0048           C=0.
ISN 0049       3   E=E*A/Z
ISN 0050           C=C+E
ISN 0051           Z=Z+1.0
ISN 0052           IF(Z.LE.X) GO TO 3
ISN 0054           CHIPRB=C*Y+S
ISN 0055           RETURN
ISN 0056           END

****** END OF COMPILATION ******

U OF A   JOB STATISTICS --   320 CARDS READ --   359 LINES PRINTED --   0 CARDS PUNCHED --   0.55 MINUTES EXECUTION TIME
```